# RATIONAL EVOLUTION OF CYTOKINES FOR HIGHER STABILITY, THE CYTOKINES AND ENCODING NUCLEIC ACID MOLECULES

## RELATED APPLICATIONS

Benefit of priority to under 37 C.F.R. §1.119(e) to U.S. provisional application Serial No. 60/457,135, entitled "RATIONAL EVOLUTION OF CYTOKINES FOR HIGHER STABILITY, ENCODING NUCLEIC ACID MOLECULES AND RELATED APPLICATIONS," filed March 21, 2003, and to U.S. provisional application Serial No. 60/409,898, entitled "RATIONAL EVOLUTION OF CYTOKINES FOR HIGHER STABILITY, ENCODING NUCLEIC ACID MOLECULES AND RELATED APPLICATIONS," filed September 9, 2002, each to Rene Gantier, Thierry Guyon, Manuel Vega and Lila Drittanti is claimed.

This application is related to U.S. application Serial No. attorney dkt. no. 37851-922PC, entitled, "RATIONAL EVOLUTION OF CYTOKINES FOR HIGHER STABILITY, THE CYTOKINES AND ENCODING NUCLEIC ACID MOLECULES," to Rene Gantier, Thierry Guyon, Manuel Vega and Lila Drittanti. This application also is related to U.S. application Serial No. Attorney docket no. 37851-923, filed the same day herewith, entitled "RATIONAL DIRECTED PROTEIN EVOLUTION USING TWO-DIMENSIONAL RATIONAL MUTAGENESIS SCANNING," to U.S. provisional application Serial No. 60/457,063, entitled "RATIONAL DIRECTED PROTEIN EVOLUTION USING TWO-DIMENSIONAL RATIONAL MUTAGENESIS SCANNING," filed March 21, 2003, and to U.S. provisional application Serial No. 60/410,258, entitled "RATIONAL DIRECTED PROTEIN EVOLUTION USING TWO-DIMENSIONAL RATIONAL MUTAGENESIS SCANNING," filed September 9, 2002, each to Rene Gantier, Thierry Guyon, Hugo Cruz Ramos, Manuel Vega and Lila Drittanti. This application also is related to co-pending U.S. application Serial No. 10/022,249, filed December 17, 2001, entitled "HIGH THROUGHPUT DIRECTED EVOLUTION BY RATIONAL MUTAGENESIS," to Manuel Vega and Lila Drittanti.

The subject matter of each of the above-noted applications and provisional applications is incorporated by reference in its entirety.

**FIELD OF INVENTION**

Modified cytokine proteins having selected modified properties compared to the unmodified or wild type proteins, and nucleic acid molecules encoding these proteins are provided. The proteins can be used for treatment and diagnosis.

**BACKGROUND**

The delivery of therapeutic proteins for clinical use is a major challenge to pharmaceutical science. Once in the blood stream, these proteins are constantly eliminated from circulation within a short time by different physiological processes, involving metabolism as well as clearance using normal pathways for protein elimination, such as (glomerular) filtration in the kidneys or proteolysis in blood. The latter is often the limiting process affecting the half-life of proteins used as therapeutic agents in per-oral administration and either intravenous or intramuscular injection. The problems associated with these routes of administration of proteins are well known and various strategies have been used in attempts to solve them.

A protein family, which has been the focus of much clinical work, and efforts to improve its administration and bio-assimilation, is the cytokine family, including the interferon family. Interferon molecules are grouped in the heterogeneous family of cytokines, originally identified on the basis of their ability to induce cellular resistance to viral infections (Díaz *et al.*, *J. Interferon Cytokine Res.*, 16:179-180, 1996). Type I interferons, referred to as interferons $\alpha/\beta$, include many members of the interferon $\alpha$ family (interferon $\alpha 1$, $\alpha 2$, $\omega$ and $\tau$) as well as interferon $\beta$. The type II interferon $\gamma$ is different from type I in its particular mechanisms that regulate its production. Whereas the production of interferons $\alpha/\beta$ is most efficiently induced in many types of cells upon viral infection, interferon-$\gamma$ is produced mainly in cells of hemopoietic

system, such as T-cells or natural killer cells, upon stimulation by antigens or cytokines, respectively. These two interferon systems are functionally non-redundant in the antiviral defense host.

Interferon $\alpha$, hereinafter "interferon alpha-2b," or "interferon $\alpha$-2b" or "IFN$\alpha$-2b," used interchangeably, has a broad spectrum of biological effects, including antiviral effects. Antiviral effects include antiproliferative and immuno-modulatory actions (Stark *et al., Annu. Rev. Biochem.*, 67:227-264, 1998). As well as eliciting strong antiviral activities in target cells, interferons $\alpha/\beta$ also activate effector cells of the innate immune system such as natural killer cells and macrophages (Pestka *et al., Annu. Rev. Biochem.*, 56:727-777, 1987; Biron *et al., Annu. Rev. Immunol.*, 17:189-220, 1999). As part of its immuno-modulatory action, interferon type I protects T-lymphocytes from apoptosis (Scheel-Toeller *et al., Eur. J. Immunol.*, 29:2603-2612, 1999; Marrack *et al., J. Exp. Med.*, 189:521-530, 1999) and growth enhancing factors (Robert *et al., Hematol. Oncol.*, 4:113-120, 1986; Morikawa *et al., J. Immunol.*, 139:761-766, 1987). The biological effects of interferons $\alpha/\beta$ are initiated upon binding to the IFN type I receptor, which results in activation of several downstream effector molecules (Hibbert and Foster, *J. Interferon Cytokine Res.*, 19:309-318, 1999).

Interferons as well as many cytokines are important therapeutics. Since naturally occurring variants have not evolved as therapeutics, they often have undesirable side-effects as well as the above-noted problems of short-half life, administration and bioavailability. Hence, there is a need to improve properties of cytokines, including interferons, for use as therapeutic agents. Therefore, among the objects herein, it is an object to provide cytokines that have improved therapeutic properties.

## SUMMARY

Provided herein are methods for directed evolution of families of proteins and resulting families of modified proteins. A family, such as the cytokine protein family, is initially identified. A property or phenotype for modification, such as resistance to proteolysis for increased stability in blood, is selected for modification. A representative member or members of the family, such as members of the interfero $\alpha$ family, such as IFN$\alpha$-2b or IFN$\alpha$-2a, or interferon $\beta$ family, is (are) selected. It is modified using any directed evolution method and protein(s) with a desired phenotype are screened and identified. In addition, the 3-dimensional structure of the protein can be mapped to topologically and spatially identify the loci that are modified to achieve the phenotypic change. 3-dimensional structures of other members of the family are generated or obtained and compared with the modified family member. Loci in the other family members that correspond on the protein to those modified in the original protein are identified and modified. The resulting proteins can be tested to confirm that they exhibit the modified phenotype.

Provided herein are methods for generating modified cytokines based on structural homology (3D scanning). These methods are based on the spatial and topological structure; they are not based on their underlying sequences of amino acid residues. The methods are used for identification of target sites for mutagenesis, particularly in families of target proteins. The targets are identified through comparison of patterns of protein backbone folding between and among structurally related proteins. The methods are exemplifed herein for cytokines. Families of the modified cytokines also are provided herein.

Any protein known or otherwise available to those of skill in the art is suitable for modification, such as optimization or improvement of a selected property, using the directed evolution methods provided herein, including cytokines (e.g., IFN$\alpha$, including IFN$\alpha$-2b and IFN$\alpha$-2a, and IFN$\beta$) or any other proteins that have already been mutated or optimized.

Provided herein are modified cytokines that exhibit increased resistance to proteolysis as assessed *in vivo* or *in vitro*. Typically the increase in resistance is a least 5%, generally 8%, 10% or more. The modified cytokines provided herein include those designed by 3D scanning using the interferon $\alpha$'s that were modified based upon 2D scanning methods herein.

Also provided herein are modified (mutant) cytokine proteins, such as variants of IFN$\beta$ and IFN$\alpha$, including IFN$\alpha$-2b and IFN$\alpha$-2a proteins and IFN$\beta$ proteins, that have altered, particularly, improved therapeutic properties, including higher stability compared to the unmodified forms. In particular, exemplary modified cytokines provided herein have increased stability, which, for example, improves their use as therapeutics. Among the modified cytokines provided herein are those that exhibits increased resistance to proteolysis compared to the unmodified cytokine. In particular, such resistance is at least 10%, 20%, 30%, 40%, 50%, 70%, 100% or more resistant to proteolysis compared to the unmodified cytokine. Also provided are cytokines that have increased anti-proliferative and/or antiviral activity and/or resistance to proteolysis compared to an unmodified cytokine.

Exemplary of the modified cytokines provided herein are modified interferons that exhibit higher stability compared to unmodified forms. Such modified interferons can be used for treating conditions in humans that are responsive to treatment with interferons, such, but are not limited to, as viral infections, cancer or tumors, undesired cell proliferation and for immuno-modulation.

Exemplary of proteins that can be modified by the 2D and 3D scanning methods provided herein are cytokines from the interferons/interleukin-10 family. This family includes, for example, interleukin-10 (IL-10; SEQ ID NO:200, interferon beta (IFN$\beta$; SEQ ID NO: 196), interferon alpha-2a (IFN$\alpha$-2a; SEQ ID NO: 182), interferon alpha-2b (IFN$\alpha$-2b; SEQ ID NO:1), and interferon gamma (IFN-$\gamma$; SEQ ID NO: 199).

The long-chain cytokine protein family includes, among others, granulocyte colony stimulating factor (G-CSF; SEQ ID NO: 210), leukemia inhibitory factor (LIF; SEQ ID NO: 213), growth hormone (hGH; SEQ ID NO: 216), ciliary neurotrophic factor (CNTF; SEQ ID NO: 212), leptin (SEQ ID NO: 211), oncostatin M (SEQ ID NO: 214), interleukin-6 (IL-6; SEQ ID NO: 217) and interleukin-12 (IL-12; SEQ ID NO: 215). The short-chain cytokine protein family includes, among others, erythropoietin (EPO; SEQ ID NO: 201), granulocyte-macrophage colony stimulating factor (GM-CSF; SEQ ID NO: 202), interleukin-2 (IL-2; SEQ ID NO: 204), interleukin-3 (IL-3; SEQ ID NO: 205), interleukin-4 (IL-4; SEQ ID NO: 207), interleukin-5 (IL-5; SEQ ID NO: 208), interleukin-13 (IL-13; SEQ ID NO: 209), Flt3 ligand (SEQ ID NO: 203) and stem cell factor (SCF; SEQ ID NO: 206). Modified forms of each that have increased resistance to proteolysis are provided. They were generated by comparison among the 3D-structures to identify residues that improve resistance to proteolysis. Pharmaceutical compositions containing each modified cytokine and uses and methods of treatment are provided.

The modified cytokines have use as therapeutics. Each cytokine has improved biological and or therapeutic activity, compared to the know activity of the unmodified cytokine. Accordingly, uses of the cytokines for treatment of cytokine-mediated diseases and diseases for which immunotherapy is employed are provided. Methods of treatment using the modified cytokines for diseases also are provided. Each cytokine has a known therapeutic use, and such use is contemplated herein. Cyokines provided herein have improved properties, such as increased bioavailability, improved stability, particularly *in vivo*, and/or greater efficacy.

## DESCRIPTION OF THE FIGURES

Figure 1(A) displays the sequence of the mature IFN$\alpha$-2b. Residues targeted by a mixture of proteases, including $\alpha$-chymotrypsin (F, L, M, W, and Y), endoproteinase Arg-C (R), endoproteinase Asp-N (D), endoproteinase Glu-C (E), endoproteinase Lys-C (K), and trypsin (K, and R), are underlined and in bold lettering

Figure 1(B) displays the structure of IFN$\alpha$-2b obtained from the NMR structure of IFN$\alpha$-2a (PDB code 1ITF) in ribbon representation. Surface residues exposed to the action of the proteases considered in FIG1A are in space filling representation.

Figure 2 depicts the "Percent Accepted Mutation" (PAM250) matrix Values given to identical residues are shown in gray squares. Highest values in the matrix are shown in black squares and correspond to the highest occurrence of substitution between two residues.

Figure 3 presents the scores obtained from PAM250 analysis for the amino acid substitutions (replacing amino acids on the vertical axis; amino acid position on the horizontal axis) aimed at introducing resistance to proteolysis into the IFN$\alpha$-2b at the protease target sequences. The two best replacing residues for each target amino acid according to the highest substitution scores are shown in black rectangles.

Figures 4(A)-4(C) provide graphs of experiments indicating the levels of protection against *in vitro* proteolysis for IFN$\alpha$-2b variants produced in mammalian cells. In Figures 4(B) and 4(C), the vertical axis indicates the relative level of non-proteolized protein and the horizontal axis indicates time in hours.

Figure 5 displays the characterization of several IFN$\alpha$-2b variants, produced in mammalian cells, treated with $\alpha$-chymotrypsin.

Figure 6(A) shows the characterization of the E113H IFN$\alpha$-2b variant when treated with $\alpha$-chymotrypsin. The percent of residual (anti-viral) activity for the variant (black line and white circles) after treatment with $\alpha$-chymotrypsin was compared to the treated wild-type IFN$\alpha$-2b

(dashed line and black squares). For this experiment, the E113H IFN$\alpha$-2b variant was produced in mammalian cells.

Figure 6(B) shows the characterization of the E113H IFN$\alpha$-2b variant treated with a mixture of proteases. The percent of residual (anti-viral) activity for the variant (black line and white circles) after treatment with protease mixture was compared to the treated wild-type IFN$\alpha$-2b (dashed line and black squares). For this experiment, the E113H IFN$\alpha$-2b variant was produced in mammalian cells.

Figure 6(C) presents the characterization of the E113H IFN$\alpha$-2b variant treated with blood lysate. The percent of residual (anti-viral) activity for the variant (black line and white circles) after treatment with blood lysate was compared to the treated wild-type IFN$\alpha$-2b (dashed line and black squares). For this experiment, the E113H IFN$\alpha$-2b variant was produced in mammalian cells.

Figure 6(D) presents the characterization of the E113H IFN$\alpha$-2b variant treated with serum. The percent of residual (anti-viral) activity for the variant (black line and white circles) after treatment with serum was compared to the treated wild-type IFN$\alpha$-2b (dashed line and black squares). For this experiment, the E113H IFN$\alpha$-2b variant was produced in mammalian cells.

Figures 6(E) and 6(F) provide graphs indicating the levels of protection against in vitro proteolysis for IFN$\alpha$-2b variants produced in bacteria. In Figures 6(E) and 6(F), the vertical axis indicates the relative level of non-proteolized protein and the horizontal axis indicates time in hours. The percent of residual (anti-viral) activity for the variants (gray circles with continuous lines) after treatment were compared to the treated wild-type IFN$\alpha$-2b (solid circles with dashed lines).

Figure 6(G) provides graphs indicating the in vitro potency for antiviral activity, for IFN$\alpha$-2b variants produced in bacteria. The vertical axis indicates the level of antiviral activity and the horizontal axis indicates concentration of the variants at which each level of activity is

achieved. The activity for the variants (continous line with gray circles) was compared to that of the wild-type IFN$\alpha$-2b (black triangles with dashed lines). The potency for each variant was calculated from the graphs as the concentration at the inflection point of the respective curves. Figure 6(T) shows the value of potency obtained for each variant tested compared to the wild type IFN$\alpha$.

Figure 6(H) provides the in vitro potency for anti-proliferation activity, for IFN$\alpha$-2b variants produced in bacteria. The activity for the variants was compared to that of the wild-type IFN$\alpha$-2b in serial dilution experiments where the anti-proliferation activity was measured for a number of dilutions for each variant. Potency was calculated from the graphs as the concentration at the inflection point of the respective curves. The figure shows the value of potency obtained for each variant tested and in comparison to the wild type IFN$\alpha$.

Figures 6(I) to 6(N) provide graphs indicating the pharmacokinetics in mice following subcutaneous injection of IFN$\alpha$-2b variants produced in bacteria. The vertical axis indicates the level of antiviral activity in blood and the horizontal axis indicates the time after injection at which the level of antiviral activity is determined. The pharmacokinetics of the variants (in gray solid circles with gray continuous lines) was compared to that of the wild-type IFN$\alpha$-2b (in black with dashed lines) and of a pegylated derivative (Pegasys, Roche) (36$\mu$g/ml open triangles with continuous black lines; and 18 $\mu$g/ml open circles with continuous black lines); and vehicle (gray solid triangles with continous gray lines. The Area Under the Curve (AUC) for each variant was calculated from the graphs and is shown in 6(U).

Figure 6(O) provides graphs indicating the levels of protection against *in vitro* proteolysis for IFN$\beta$ variants produced in mammalian cells.

Figure 6(N), the vertical axis indicates the relative level of non-proteolized protein and the horizontal axis indicates time in hours. The

percent of residual (anti-viral) activity for the variants after treatment were compared to the treated wild-type IFN$\beta$.

Figures 6(P) to 6(S) provide graphs indicating the in vitro potency for either antiviral activity (6(P) and 6(Q)) or anti-proliferative activity (6(R) and 6(S), for a number of IFN$\beta$ variants produced in mammalian cells. The vertical axis indicates the level of (antiviral or anti-proliferation) activity and the horizontal axis indicates the concentration of the variants at which each level of activity is achieved. The activity for the variants (6(Q) and (6(S)) was compared to that of the wild-type IFN$\beta$ (6(P) and (6(R)). The activity obtained with either no previous treatment or by treating the variants with proteases prior to the activity test is shown.

Figure 6(T) provides a comparison of antiviral activity (potency), anti-proliferation activity (potency), number of mutations present and AUC (from PK) for a number of IFN$\alpha$-2b and in comparison with the wild-type IFN$\alpha$-2b.

Figure 6(U) provides IFN units injected and protein injected ($\mu$g/ml) for the data in Figure 6(T).

Figure 7(A) depicts a top view ribbon representation of IFN$\alpha$-2b structure obtained from the NMR structure of IFN$\alpha$-2$\alpha$ (PDB code 1ITF). Residues represented in "space filling" define (1) the "receptor binding region" based on either our "alanine scanning" analysis or on studies by Piehler *et al.*, *J. Biol. Chem.*, 275:40425-40433, 2000, and Roisman *et al.*, *Proc. Natl. Acad. Sci. USA*, 98:13231-13236, 2001 (in light-gray and dark-gray, respectively), and (2) replacing residues (LEADs) for resistance to proteolysis (in black).

Figure 7(B) depicts a side view ribbon representation of IFN$\alpha$-2b structure (PDB code 1ITF). Residue representation is as in FIG7A.

Figure 8(A) schematizes the identification of homologous amino acid positions between a number of cytokines and the LEAD mutants of IFN$\alpha$-2b using 3-dimensional scanning (also referred to herein as based on "structure-based homology" methods or "structural homology" methods).

Figure 8(B) illustrates a structural overlapping between human interferon $\alpha$-2b obtained from the NMR structure of IFN$\alpha$-2a (PDB code 1ITF) and human interferon $\beta$ (PDB code 1AU1) using Swiss Pdb Viewer.

Figure 8(C) illustrates a structural overlapping between human interferon $\alpha$-2b obtained from the NMR structure of IFN$\alpha$-2a (PDB code 1ITF) and erythropoietin (PDB code 1BUY) using Swiss Pdb Viewer.

Figure 8(D) illustrates a structural overlapping between human interferon $\alpha$-2b obtained from the NMR structure of IFN$\alpha$-2a (PDB code 1ITF) and granulocyte-colony stimulating factor (PDB code 1CD9) using Swiss Pdb Viewer.

Figure 9 illustrates a structural alignment of a number of cytokines and interferon $\alpha$-2b sequences. Bold underlined residues define the region on each cytokine sequence that based on structural homology comparison corresponds to the structurally-related mutations found on the LEADs for protease resistance of IFN$\alpha$-2b.

Figure 10(A) shows the antiviral activity of interferon $\alpha$-2b mutants generated by alanine-scanning analysis used for protein redesign. Plotted symbols for wild type and variants of interferon $\alpha$-2b are indicated in the inset.

Figure 10(B) displays cell proliferation after treatment with interferon $\alpha$-2b mutants obtained by alanine-scanning analysis. Plotted symbols for wild type and variants of interferon $\alpha$-2b are indicated in the inset.

Figure 10(C) displays the correlation between the antiviral activity and cell proliferation activity of interferon $\alpha$-2b mutants obtained by alanine-scanning analysis.

Figure 11 Candidate glycosylation sites for interferon $\alpha$-2b stabilization and redesign thereof.

Figure 12 (A) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of

interferon $\beta$ (corresponding to SEQ ID Nos: 233-289), based on 3D-scanning (structural homology method), including PAM250 analysis.

Figure 12 (B) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interferon gamma (corresponding to SEQ ID Nos: 290-311), based on structural homology and PAM250 analysis.

Figure 12 (C) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-10 (corresponding to SEQ ID Nos: 312-361), based on structural homology and PAM250 analysis.

Figure 12 (D) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of ciliary neurotrophic factor (corresponding to SEQ ID Nos: 684-728), based on structural homology and PAM250 analysis.

Figure 12 (E) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of granulocyte-colony stimulating factor (corresponding to SEQ ID Nos: 631-662), based on structural homology and PAM250 analysis.

Figure 12 (F) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of human growth hormone (corresponding to SEQ ID Nos: 850-895), based on structural homology and PAM250 analysis.

Figure 12 (G) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-12 (corresponding to SEQ ID Nos: 794-849), based on structural homology and PAM250 analysis.

Figure 12 (H) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-6 (corresponding to SEQ ID Nos: 896-939), based on structural homology and PAM250 analysis.

Figure 12 (I) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of leptin (corresponding to SEQ ID Nos: 663-683), based on structural homology and PAM250 analysis.

Figure 12 (J) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of leukemia inhibitory factor (corresponding to SEQ ID Nos: 729-760), based on structural homology and PAM250 analysis.

Figure 12 (K) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of oncostatin M (corresponding to SEQ ID Nos: 761-793), based on structural homology and PAM250 analysis.

Figure 12 (L) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of erythropoietin (corresponding to SEQ ID Nos: 940-977), based on structural homology and PAM250 analysis.

Figure 12 (M) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of Flt3 ligand (corresponding to SEQ ID Nos: 401-428), based on structural homology and PAM250 analysis.

Figure 12 (N) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of granulocyte-macrophage colony-stimulating factor (corresponding to SEQ ID Nos: 362-400), based on structural homology and PAM250 analysis.

Figure 12 (O) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-13 (corresponding to SEQ ID Nos: 603-630), based on structural homology and PAM250 analysis.

Figure 12 (P) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of

interleukin-2 (corresponding to SEQ ID Nos: 429-476), based on structural homology and PAM250 analysis.

Figure 12 (Q) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-3 (corresponding to SEQ ID Nos: 477-498), based on structural homology and PAM250 analysis.

Figure 12 (R) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-4 (corresponding to SEQ ID Nos: 543-567), based on structural homology and PAM250 analysis.

Figure 12 (S) shows the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of interleukin-5 (corresponding to SEQ ID Nos: 568-602), based on structural homology and PAM250 analysis.

Figure 12 (T) displays the is-HIT residue positions and type of replacing amino acids selected to generate modified protein sequences of stem cell factor (corresponding to SEQ ID Nos: 499-542), based on structural homology and PAM250 analysis.

## DETAILED DESCRIPTION

**OUTLINE**

- A. Definitions
- B. Directed Evolution
  - 1) Pure Random Mutagenesis
  - 2) Restricted Random Mutagenesis
  - 3) Non-restricted Rational mutagenesis
- C. 2-Dimensional Rational Scanning (2D scanning)
  - 1) Identifying In-silico HITS
  - 2) Identifying Replacing Amino Acids
    - (a) Percent Accepted Mutation (PAM)
      - (i) PAM Analysis
      - (ii) PAM250
    - (b) Jones and Gonnet
    - (c) Fitch and Feng
    - (d) McLachlan, Grantham and Miyata
    - (e) Rao
    - (f) Risler
    - (g) Johnson
    - (h) Block Substitution Matrix (BLOSUM)
  - 3) Physical Construction of Mutant Proteins and Biological Assays

## A. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference in their entirety. In the event that there is a plurality of definitions for terms herein, those in this section prevail. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

As used herein, biological activity of a protein refers to any activity manifested by the protein *in vivo*.

As used herein, "a directed evolution method" refers to methods that "adapt" either natural proteins, synthetic proteins or protein domains to work in new or existing natural or artificial chemical or biological environments and/or to elicit new functions and/or to increase or decrease a given activity, and/or to modulate a given feature. Exemplary directed evolution methods include pure random mutageneis methods; restricted random mutagenesis methods; and non-restricted rational mutagenesis methods, such as the rational directed evolution method described in co-pending U.S. application Serial No. 10/022,249; and the 2-dimensional rational scanning method provided herein.

As used herein, two dimensional rational mutagenesis scanning (2D scanning) refers to the processes provided herein in which two dimensions of a particular protein sequence are scanned: (1) one dimension is to identify specific amino acid residues along the protein sequence to replace with different amino acids, referred to as is-HIT target positions, and (2) the second dimension is the amino acid type selected for replacing the particular is-HIT target, referred to as the replacing amino acid.

As used herein, *in silico* refers to research and experiments performed using a computer. In silico methods include, but are not limited to, molecular modeling studies, and biomolecular docking experiments.

As used herein, "is-HIT" refers to an *in silico* identified amino acid position along a target protein sequence that has been identified based on *i*) the particular protein properties to be evolved, *ii*) the protein's amino acid sequence, and/or *iii*) the known properties of the individual amino acids. These is-HIT loci on the protein sequence are identified without use of experimental biological methods. For example, once the protein feature(s) to be optimized is (are) selected, diverse sources of information or previous knowledge (i.e., protein primary, secondary or tertiary structures, literature, patents) are exploited to determine those amino acid

positions that may be amenable to improved protein fitness by replacement with a different amino acid. This step utilizes protein analysis "*in silico*." All possible candidate amino acid positions along a target proteins primary sequence that might be involved in the feature being evolved are referred to herein as "*in silico* HITs" ("is-HITs"). The collection (library), of all is-HITs identified during this step represents the first dimension (target residue position) of the two-dimensional scanning methods provided herein.

As used herein, "amenable to providing the evolved predetermined property or activity," in the context of identifying is-HITs, refers to an amino acid position on a protein that is contemplated, based on *in silico* analysis, to possess properties or features that when replaced would result in the desired activity being evolved. The phrase "amenable to providing the evolved predetermined property or activity," in the context of identifying replacement amino acids, refers to a particular amino acid type that is contemplated, based on *in silico* analysis, to possess properties or features that when used to replace the original amino acid in the unmodified starting protein would result in the desired activity being evolved.

As used herein, high-throughput screening (HTS) refers to processes that test a large number of samples, such as samples of test proteins or cells containing nucleic acids encoding the proteins of interest to identify structures of interest or the identify test compounds that interact with the variant proteins or cells containing them. HTS operations are amenable to automation and are typically computerized to handle sample preparation, assay procedures and the subsequent processing of large volumes of data.

As used herein, the term "restricted," when used in the context of the identification of is-HIT amino acid positions along the protein sequence selected for amino acid replacement and/or the identification of replacing amino acids, means that fewer than all amino acids on the

protein-backbone are selected for amino acid replacement; and/or fewer than all of the remaining 19 amino acids available to replace the original amino acid present in the unmodified starting protein are selected for replacement. In particular embodiments of the methods provided herein, the is-HIT amino acid positions are restricted, such that fewer than all amino acids on the protein-backbone are selected for amino acid replacement. In other embodiments, the replacing amino acids are restricted, such that fewer than all of the remaining 19 amino acids available to replace the native amino acid present in the unmodified starting protein are selected as replacing amino acids. In a particular embodiment, both of the scans to identify is-HIT amino acid positions and the replacing amino acids are restricted, such that fewer than all amino acids on the protein-backbone are selected for amino acid replacement and fewer than all of the remaining 19 amino acids available to replace the native amino acid are selected for replacement.

As used herein, "candidate LEADs," are mutant proteins that are contemplated as potentially having an alteration in any attribute, chemical, physical or biological property in which such alteration is sought. In the methods herein, candidate LEADs are generally generated by systematically replacing is-HITS loci in a protein or a domain thereof with typically a restricted subset, or all, of the remaining 19 amino acids, such as obtained using PAM analysis. Candidate LEADs can be generated by other methods known to those of skill in the art tested by the high throughput methods herein.

As used herein, "LEADs" are "candidate LEADs" whose activity has been demonstrated to be optimized or improved for the particular attribute, chemical, physical or biological property. For purposes herein a "LEAD" typically has activity with respect to the function of interest that differs by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200% or more from the unmodified and/or wild type (native) protein. In certain embodiments, the change in activity is at least

about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100%, of the activity of the unmodified target protein. In other embodiments, the change in activity is not more than about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100%, of the activity of the unmodified target protein. In yet other embodiments, the change in activity is at least about 2 times, 3 times, 4 times, 5 times, 6 times, 7 times, 8 times, 9 times, 10 times, 20 times, 30 times, 40 times, 50 times, 60 times, 70 times, 80 times, 90 times, 100 times, 200 times, 300 times, 400 times, 500 times, 600 times, 700 times, 800 times, 900 times, 1000 times, or more greater than the activity of the unmodified target protein. The desired alteration, which can be either an increase or a reduction in activity, will depend upon the function or property of interest (e.g., ±10%, ±20%, etc.). The LEADs may be further optimized by replacement of a plurality (2 or more) of "is-HIT" target positions on the same protein molecule to generate "super-LEADs."

As used herein, the term "super-LEAD" refers to protein mutants (variants) obtained by combining the single mutations present in two or more of the LEAD molecules into a single protein molecule. Accordingly, in the context of the modified proteins provided herein, the phrase "proteins comprising one or more single amino acid replacements" encompasses any combination of two or more of the mutations described herein for a respective protein. For example, the modified proteins provided herein having one or more single amino acid replacements can have can have any combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more of the amino acid replacements at the disclosed replacement positions. The collection of super-LEAD mutant molecules is generated, tested and phenotypically characterized one-by-one in addressable arrays. Super-LEAD mutant molecules are such that each molecule contains a variable number and type of LEAD mutations. Those molecules displaying further improved fitness for the particular feature being evolved, are referred to as super-LEADs. Super-LEADs can

be generated by other methods known to those of skill in the art and tested by the high throughput methods herein. For purposes herein a super-LEAD typically has activity with respect to the function of interest that differs from the improved activity of a LEAD by a desired amount, such as at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200% or more from at least one of the LEAD mutants from which it is derived. As with LEADs, the change in the activity for super-LEADs is dependent upon the activity that is being "evolved." The desired alteration, which can be either an increase or a reduction in activity, will depend upon the function or property of interest.

As used herein, a recitation that modified protein has more antiviral activity (or other activity) than antiproliferative activity (or another activity) compared to the unmodified cytokine, is comparing the absolute value of the change in each activity compared to wild type.

As used herein, the phrase "altered loci" refers to the is-HIT amino acid positions in the LEADs or super-LEADs that are replaced with different replacing amino acids, resulting in the desired altered phenotype or activity.

As used herein, an exposed residue presents more than 15% of its surface exposed to the solvent.

As used herein, the phrase "structural homology" refers to the degree of coincidence in space between two or more protein backbones. Protein backbones that adopt the same protein structure, fold and show similarity upon three-dimensional structural superposition in space can be considered structurally homologous. Structural homology is not based on sequence homology, but rather on three-dimension homology. Two amino acids in two different proteins said to be homologous based on structural homology between those proteins, do not necessarily need to be in sequence-based homologous regions. For example, protein backbones that have a root mean squared (RMS) deviation of less than 3.5, 3.0, 2.5, 2.0, 1.7 or 1.5 angstrom at a given space position or

defined region between each other can be considered to be structurally homologous in that region, and are referred to herein as having a "high coincidence" between their backbones. It is contemplated herein that substantially equivalent (e.g., "structurally related") amino acid positions that are located on two or more different protein sequences that share a certain degree of structural homology will have comparable functional tasks; also referred to herein as "structurally homologous loci." These two amino acids than can be said to be "structurally similar" or "structurally related" with each other, even if their precise primary linear positions on the amino acid sequences, when these sequences are aligned, do not match with each other. Amino acids that are "structurally related" can be far away from each other in the primary protein sequences, when these sequences are aligned following the rules of classical sequence homology.

As used herein, a structural homolog is a protein that is generated by structural homology.

As used herein, the phrase "unmodified target protein," "unmodified protein" or "unmodified cytokine," or grammatical variations thereof, refers to a starting protein that is selected for modification using the methods provided herein. The starting unmodified target protein can be the naturally occurring, wild type form of a protein. In addition, the starting unmodified target protein may have previously been altered or mutated, such that it differs from the native wild type isoform, but is nonetheless referred to herein as an starting unmodified target protein relative to the subsequently modified proteins produced herein. Thus, existing proteins known in the art that have previously been modified to have a desired increase or decrease in a particular biological activity compared to an unmodified reference protein can be selected and used herein as the starting "unmodified target protein." For example, a protein that has been modified from its native form by one or more single amino acid changes and possesses either an increase or decrease in a desired

activity, such as resistance to proteolysis, can be utilized with the methods provided herein as the starting unmodified target protein for further modification of either the same or a different biological activity.

Likewise, existing proteins known in the art that have previously been modified to have a desired increase or decrease in a particular biological activity compared to an unmodified reference protein can be selected and used herein for identification of structurally homologous loci on other structurally homologous target proteins. For example, a protein that has been modified by one or more single amino acid changes and possesses either an increase or decrease in a desired activity, such as resistance to proteolysis, can be utilized with the methods provided herein to identify on structurally homologous target proteins, corresponding structurally homologous loci that can be replaced with suitabale replacing amino acids and tested for either an increase or decrease in the desired biological actiity.

As used herein, the phrase "only one amino acid replacement occurs on each target protein" refers to the modification of a target protein, such that it differs from the unmodified form of the target protein by a single amino acid change. For example, in one embodiment, mutagenesis is performed by the replacement of a single amino acid residue at only one is-HIT target position on the protein backbone (e.g., "one-by-one" in addressable arrays), such that each individual mutant generated is the single product of each single mutagenesis reaction. The single amino acid replacement mutagenesis reactions are repeated for each of the replacing amino acids selected at each of the is-HIT target positions. Thus, a plurality of mutant protein molecules are produced, whereby each mutant protein contains a single amino acid replacement at only one of the is-HIT target positions.

As used herein, the phrase "pseudo-wild type," in the context of single or multiple amino acid replacements, are those amino acids that, while different from the original, such as native, amino acid at a given

amino acid position, can replace the native one at that position without introducing any measurable change in a particular protein activity. A population of sets of nucleic acid molecules encoding a collection of mutant molecules is generated and phenotypically characterized such that proteins with amino acid sequences different from the original amino acid, but that still elicit substantially the same level (i.e., at least 10%, 50%, 70%, 90%, 95%, 100%, depending upon the protein) and type of desired activity as the original protein are selected.

As used herein, biological and pharmacological activity includes any activity of a biological pharmaceutical agent and includes, but is not limited to, resistance to proteolysis, biological efficiency, transduction efficiency, gene/transgene expression, differential gene expression and induction activity, titer, progeny productivity, toxicity, cytotoxicity, immunogenicity, cell proliferation and/or differentiation activity, anti-viral activity, morphogenetic activity, teratogenetic activity, pathogenetic activity, therapeutic activity, tumor suppressor activity, ontogenetic activity, oncogenetic activity, enzymatic activity, pharmacological activity, cell/tissue tropism and delivery.

As used herein, a "small region" on a polypeptide is relative term depending upon the size of the polypeptide, but typically refers to a region that is less than about 10%, 15%, 25% of the protein. A large region is greater than about 10%, 15% or 25% of the protein.

As used herein, "output signal" refers to parameters that can be followed over time and, if desired, quantified. For example, when a recombinant protein is introduced into a cell, the cell containing the recombinant protein undergoes a number of changes. Any such change that can be monitored and used to assess the transformation or transfection, is an output signal, and the cell is referred to as a reporter cell; the encoding nucleic acid is referred to as a reporter gene, and the construct that includes the encoding nucleic acid is a reporter construct. Output signals include, but are not limited to, enzyme activity,

fluorescence, luminescence, amount of product produced and other such signals. Output signals include expression of a gene or gene product, including heterologous genes (transgenes) inserted into the plasmid virus. Output signals are a function of time ("t") and are related to the amount of protein used in the composition. For higher concentrations of protein, the output signal can be higher or lower. For any particular concentration, the output signal increases as a function of time until a plateau is reached. Output signals can also measure the interaction between cells, expressing heterologous genes, and biological agents

As used herein, the activity of an IFN$\alpha$-2b or IFN$\alpha$-2a protein refers to any biological activity that can be assessed. In particular, herein, the activity assessed for the IFN$\alpha$-2b or IFN$\alpha$-2a proteins is resistance to proteolysis, antiviral activity and cell proliferation activity.

As used herein, the Hill equation is a mathematical model that relates the concentration of a drug (*i.e.*, test compound or substance) to the response measured

$$y = \frac{y_{max}[D]^n}{[D]^n + [D_{50}]^n} ,$$

where y is the variable measured, such as a response, signal, $y_{max}$ is the maximal response achievable, [D] is the molar concentration of a drug, $[D_{50}]$ is the concentration that produces a 50% maximal response to the drug, n is the slope parameter, which is 1 if the drug binds to a single site and with no cooperativity between or among sites. A Hill plot is $\log_{10}$ of the ratio of ligand-occupied receptor to free receptor vs. log [D] (M). The slope is n, where a slope of greater than 1 indicates cooperativity among binding sites, and a slope of less than 1 can indicate heterogeneity of binding. This general equation has been employed for assessing interactions in complex biological systems (see, published International PCT application No. WO 01/44809 based on PCT No. PCT/FR00/03503, see also, the EXAMPLES).

As used herein, in the Hill-based analysis (published International PCT application No. WO 01/44809 based on PCT No. PCT/FR00/03503), the parameters, $\pi,\kappa,\tau,\epsilon,\eta,\theta$, are as follows:

$\pi$ is the potency of the biological agent acting on the assay (cell-based) system;

$\kappa$ is the constant of resistance of the assay system to elicit a response to a biological agent;

$\epsilon$ is the global efficiency of the process or reaction triggered by the biological agent on the assay system;

$\tau$ is the apparent titer of the biological agent;

$\theta$ is the absolute titer of the biological agent; and

$\eta$ is the heterogeneity of the biological process or reaction.

In particular, as used herein, the parameters $\pi$ (potency) or $\kappa$ (constant of resistance) are used to respectively assess the potency of a test agent to produce a response in an assay system and the resistance of the assay system to respond to the agent.

As used herein, $\epsilon$ (efficiency), is the slope at the inflexion point of the Hill curve (or, in general, of any other sigmoidal or linear approximation), to assess the efficiency of the global reaction (the biological agent and the assay system taken together) to elicit the biological or pharmacological response.

As used herein, $\tau$ (apparent titer) is used to measure the limiting dilution or the apparent titer of the biological agent.

As used herein, $\theta$ (absolute titer), is used to measure the absolute limiting dilution or titer of the biological agent.

As used herein, $\eta$ (heterogeneity) measures the existence of discontinuous phases along the global reaction, which is reflected by an abrupt change in the value of the Hill coefficient or in the constant of resistance.

As used herein, a population of sets of nucleic acid molecules encoding a collection (library) of mutants refers to a collection of plasmids

or other vehicles that carry (encode) the gene variants, such that individual plasmids or other individual vehicles carry individual gene variants. Each element (member) of the collection is physically separated from the others, such as individually in an appropriate addressable array, and has been generated as the single product of an independent mutagenesis reaction. When a collection (library) of such proteins is contemplated, it will be so-stated.

As used herein, a "reporter cell" is the cell that "reports," *i.e.*, undergoes the change, in response to a condition, such as, for example, exposure to a protein or a virus or to a change it its external or internal environment.

As used herein, "reporter" or "reporter moiety" refers to any moiety that allows for the detection of a molecule of interest, such as a protein expressed by a cell. Reporter moieties include, but are not limited to, for example, fluorescent proteins, such as red, blue and green fluorescent proteins; LacZ and other detectable proteins and gene products. For expression in cells, nucleic acid encoding the reporter moiety can be expressed as a fusion protein with a protein of interest or under to the control of a promoter of interest.

As used herein, phenotype refers to the physical, physiological or other manifestation of a genotype (a sequence of a gene). In methods herein, phenotypes that result from alteration of a genotype are assessed.

As used herein, "activity" means in the largest sense of the term any change in a system (either biological, chemical or physical system) of any nature (changes in the amount of product in an enzymatic reaction, changes in cell proliferation, in immunogenicity, in toxicity) caused by a protein or protein mutant when they interact with that system. In addition, the term "activity," "higher activity" or "lower activity" as used herein in reference to resistance to proteases, proteolysis, incubation with serum or with blood, means the ratio or residual biological (antiviral)

activity between "after" protease/blood or serum treatment and "before" protease/blood or serum treatment.

As used herein, activity refers to the function or property to be evolved. An active site refers to a site(s) responsible or that participates in conferring the activity or function. The activity or active site evolved (the function or property and the site conferring or participating in conferring the activity) can have nothing to do with natural activities of a protein. For example, it could be an 'active site' for conferring immunogenicity (immunogenic sites or epitopes) on a protein.

As used herein, treatment means any manner in which the symptoms of a condition, disorder or disease are ameliorated or otherwise beneficially altered. Treatment also encompasses any pharmaceutical use of the modified cytokines and compositions provided herein herein.

As used herein, cytokine-mediated or cytokine-involved diseases refer to diseases in which cytokines potentiate, cause or are involved in the disease process or to diseases in which administration of a cytokine is ameliorative of a disease or symptoms thereof. Cytokines can be used in immunotherapeutic therapies or protocols.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their known, three-letter or one-letter abbreviations (see, Table 1). The nucleotides, which occur in the various nucleic acid fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, amino acid residue refers to an amino acid formed upon chemical digestion (hydrolysis) of a polypeptide at its peptide linkages. The amino acid residues described herein are presumed to be in the "L" isomeric form. Residues in the "D" isomeric form, which are so-designated, can be substituted for any L-amino acid residue, as long as the desired functional property is retained by the polypeptide. $NH_2$ refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxyl terminus

of a polypeptide. In keeping with standard polypeptide nomenclature described in *J. Biol. Chem.*, 243:3552-3559, 1969, and adopted at 37 C.F.R. §§ 1.821 - 1.822, abbreviations for amino acid residues are shown in Table 1:

**Table 1**
Table of Correspondence

| SYMBOL | | |
|---|---|---|
| **1-Letter** | **3-Letter** | **AMINO ACID** |
| Y | Tyr | tyrosine |
| G | Gly | glycine |
| F | Phe | phenylalanine |
| M | Met | methionine |
| A | Ala | alanine |
| S | Ser | serine |
| I | Ile | isoleucine |
| L | Leu | leucine |
| T | Thr | threonine |
| V | Val | valine |
| P | Pro | proline |
| K | Lys | lysine |
| H | His | histidine |
| Q | Gln | glutamine |
| E | Glu | glutamic acid |
| Z | Glx | Glu and/or Gln |
| W | Trp | tryptophan |
| R | Arg | arginine |
| D | Asp | aspartic acid |
| N | Asn | asparagine |
| B | Asx | Asn and/or Asp |

| SYMBOL | | |
|---|---|---|
| C | Cys | cysteine |
| X | Xaa | Unknown or other |

It should be noted that all amino acid residue sequences represented herein by formulae have a left to right orientation in the conventional direction of amino-terminus to carboxyl-terminus. In addition, the phrase "amino acid residue" is broadly defined to include the amino acids listed in the Table of Correspondence (Table 1) and modified and unusual amino acids, such as those referred to in 37 C.F.R. §§ 1.821-1.822, and incorporated herein by reference. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino acid residues or to an amino-terminal group such as $NH_2$ or to a carboxyl-terminal group such as COOH.

As used herein, nucleic acids include DNA, RNA and analogs thereof, including protein nucleic acids (PNA) and mixture thereof. Nucleic acids can be single or double stranded. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique of low copy number (typically less than 5, generally less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous of sequence complementary to or identical a gene of interest. Probes and primers can be 10, 14, 16, 20, 30, 50, 100 or more nucleic acid bases long.

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide can be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, "corresponding structurally-related" positions on two or more proteins, such as the IFN$\alpha$-2b protein and other cytokines, refers those amino acid positions determined based upon structural homology to maximize tri-dimensional overlapping between proteins.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 100% relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polypeptide length of 100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differ from that of the reference polypeptides. Similar comparisons can be made between a test and reference polynucleotides. Such differences can be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they can be clustered in one or more locations of varying length up to the maximum allowable, e.g., 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

As used herein, the phrase "sequence-related proteins" refers to proteins that have at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% amino acid identity or homology with each other.

As used herein, families of non-related proteins or "sequence-non-related proteins" refers to proteins that have less than 50%, less than 40%, less than 0%, less thant 20% amino acid identity or homology with each other.

As used herein, it also is understood that the terms "substantially identical" or "similar" varies with the context as understood by those skilled in the relevant art.

As used herein, heterologous or foreign nucleic acid, such as DNA and RNA, are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which

is found in a location or locations in the genome that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not normally produced by the cell in which it is expressed. Heterologous DNA herein encompasses any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed. Heterologous DNA and RNA can also encode RNA or proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes. Examples of heterologous nucleic acid include, but are not limited to, nucleic acid that encodes traceable marker proteins, such as a protein that confers drug resistance, nucleic acid that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and DNA that encodes other types of proteins, such as antibodies.

Hence, herein heterologous DNA or foreign DNA, includes a DNA molecule not present in the exact orientation and position as the counterpart DNA molecule found in the genome. It can also refer to a DNA molecule from another organism or species (*i.e.*, exogenous).

As used herein, a therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of disease.

As used herein, isolated with reference to a nucleic acid molecule or polypeptide or other biomolecule means that the nucleic acid or polypeptide has separated from the genetic environment from which the polypeptide or nucleic acid were obtained. It can also mean altered from the natural state. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is

"isolated," as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of a compound can be substantially purified by the one-step method described in Smith *et al.*, *Gene*, 67:31-40, 1988. The terms isolated and purified are sometimes used interchangeably.

Thus, by "isolated" is meant that the nucleic is free of the coding sequences of those genes that, in the naturally-occurring genome of the organism (if any) immediately flank the gene encoding the nucleic acid of interest. Isolated DNA can be single-stranded or double-stranded, and can be genomic DNA, cDNA, recombinant hybrid DNA, or synthetic DNA. It can be identical to a starting DNA sequence, or can differ from such sequence by the deletion, addition, or substitution of one or more nucleotides.

Isolated or purified as it refers to preparations made from biological cells or hosts means any cell extract containing the indicated DNA or protein including a crude extract of the DNA or protein of interest. For example, in the case of a protein, a purified preparation can be obtained following an individual technique or a series of preparative or biochemical techniques and the DNA or protein of interest can be present at various degrees of purity in these preparations. The procedures can include for example, but are not limited to, ammonium sulfate fractionation, gel filtration, ion exchange change chromatography, affinity chromatography, density gradient centrifugation and electrophoresis.

A preparation of DNA or protein that is "substantially pure" or "isolated" should be understood to mean a preparation free from naturally occurring materials with which such DNA or protein is normally associated in nature. "Essentially pure" should be understood to mean a

"highly" purified preparation that contains at least 95% of the DNA or protein of interest.

A cell extract that contains the DNA or protein of interest should be understood to mean a homogenate preparation or cell-free preparation obtained from cells that express the protein or contain the DNA of interest. The term "cell extract" is intended to include culture media, especially spent culture media from which the cells have been removed.

As used herein, "a targeting agent" refers to any molecule that can bind another target-molecule, such as an antibody, receptor, or ligand.

As used herein, receptor refers to a biologically active molecule that specifically binds to (or with) other molecules. The term "receptor protein" can be used to more specifically indicate the proteinaceous nature of a specific receptor.

As used herein, recombinant refers to any progeny formed as the result of genetic engineering.

As used herein, a promoter region refers to the portion of DNA of a gene that controls transcription of the DNA to which it is operatively linked. The promoter region includes specific sequences of DNA that are sufficient for RNA polymerase recognition, binding and transcription initiation. This portion of the promoter region is referred to as the promoter. In addition, the promoter region includes sequences that modulate this recognition, binding and transcription initiation activity of the RNA polymerase. These sequences can be *cis* acting or can be responsive to *trans* acting factors. Promoters, depending upon the nature of the regulation, can be constitutive or regulated.

As used herein, the phrase "operatively linked" generally means the sequences or segments have been covalently joined into one piece of DNA, whether in single or double stranded form, whereby control or regulatory sequences on one segment control or permit expression or replication or other such control of other segments. The two segments are not necessarily contiguous. For gene expression a DNA sequence and

a regulatory sequence(s) are connected in such a way to control or permit gene expression when the appropriate molecular, e.g., transcriptional activator proteins, are bound to the regulatory sequence(s).

As used herein, production by recombinant means by using recombinant DNA methods means the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA, including cloning expression of genes and methods, such as gene shuffling and phage display with screening for desired specificities.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results in more than one type of mRNA.

As used herein, a composition refers to any mixture of two or more products or compounds. It can be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, a combination refers to any association between two or more items.

As used herein, substantially identical to a product means sufficiently similar so that the property of interest is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of exemplary vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Exemplary vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors." In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. "Plasmid" and "vector" are used

interchangeably as the plasmid is the most commonly used form of vector. Other such other forms of expression vectors that serve equivalent functions and that become known in the art subsequently hereto.

As used herein, vector also is used interchangeable with "virus vector" or "viral vector. In this case, which will be clear from the context, the "vector" is not self-replicating. Viral vectors are engineered viruses that are operatively linked to exogenous genes to transfer (as vehicles or shuttles) the exogenous genes into cells.

As used herein, transduction refers to the process of gene transfer into and expression in mammalian and other cells mediated by viruses. Transfection refers to the process when mediated by plasmids.

As used herein, transformation refers to the process of gene transfer into and expression in bacterial cells mediated by plasmids.

As used herein, "allele," which is used interchangeably herein with "allelic variant" refers to alternative forms of a gene or portions thereof. Alleles occupy the same locus or position on homologous chromosomes. When a subject has two identical alleles of a gene, the subject is said to be homozygous for the gene or allele. When a subject has two different alleles of a gene, the subject is said to be heterozygous for the gene. Alleles of a specific gene can differ from each other in a single nucleotide, or several nucleotides, and can include substitutions, deletions, and insertions of nucleotides. An allele of a gene also can be a form of a gene containing a mutation.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid molecule comprising an open reading frame and including at least one exon and (optionally) an intron sequence. A gene can be either RNA or DNA. Genes can include regions preceding and following the coding region (leader and trailer).

As used herein, "intron" refers to a DNA sequence present in a given gene which is spliced out during mRNA maturation.

As used herein, "nucleotide sequence complementary to the nucleotide sequence set forth in SEQ ID NO:" refers to the nucleotide sequence of the complementary strand of a nucleic acid strand having the particular SEQ ID NO:. The term "complementary strand" is used herein interchangeably with the term "complement." The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand. When referring to double stranded nucleic acids, the complement of a nucleic acid having a particular SEQ ID NO: refers to the complementary strand of the strand set forth in the particular SEQ ID NO: or to any nucleic acid having the nucleotide sequence of the complementary strand of the particular SEQ ID NO:. When referring to a single stranded nucleic acid having a nucleotide sequence corresponding to a particular SEQ ID NO:, the complement of this nucleic acid is a nucleic acid having a nucleotide sequence which is complementary to that of the particular SEQ ID NO:.

As used herein, the term "coding sequence" refers to that portion of a gene that encodes an amino acid sequence of a protein.

As used herein, the term "sense strand" refers to that strand of a double-stranded nucleic acid molecule that has the sequence of the mRNA that encodes the amino acid sequence encoded by the double-stranded nucleic acid molecule.

As used herein, the term "antisense strand" refers to that strand of a double-stranded nucleic acid molecule that is the complement of the sequence of the mRNA that encodes the amino acid sequence encoded by the double-stranded nucleic acid molecule.

As used herein, an array refers to a collection of elements, such as nucleic acid molecules, containing three or more members. An addressable array is one in which the members of the array are identifiable, typically by position on a solid phase support or by virtue of an identifiable or detectable label, such as by color, fluorescence, electronic signal (*i.e.*, RF, microwave or other frequency that does not

substantially alter the interaction of the molecules of interest), bar code or other symbology, chemical or other such label. In certain embodiments, the members of the array are immobilized to discrete identifiable loci on the surface of a solid phase or directly or indirectly linked to or otherwise associated with the identifiable label, such as affixed to a microsphere or other particulate support (herein referred to as beads) and suspended in solution or spread out on a surface.

As used herein, a support (also referred to as a matrix support, a matrix, an insoluble support or solid support) refers to any solid or semisolid or insoluble support to which a molecule of interest, typically a biological molecule, organic molecule or biospecific ligand is linked or contacted. Such materials include any materials that are used as affinity matrices or supports for chemical and biological molecule syntheses and analyses, such as, but are not limited to: polystyrene, polycarbonate, polypropylene, nylon, glass, dextran, chitin, sand, pumice, agarose, polysaccharides, dendrimers, buckyballs, polyacryl-amide, silicon, rubber, and other materials used as supports for solid phase syntheses, affinity separations and purifications, hybridization reactions, immunoassays and other such applications. The matrix herein can be particulate or can be in the form of a continuous surface, such as a microtiter dish or well, a glass slide, a silicon chip, a nitrocellulose sheet, nylon mesh, or other such materials. When particulate, typically the particles have at least one dimension in the 5-10 mm range or smaller. Such particles, referred collectively herein as "beads," are often, but not necessarily, spherical. Such reference, however, does not constrain the geometry of the matrix, which can be any shape, including random shapes, needles, fibers, and elongated. Roughly spherical "beads," particularly microspheres that can be used in the liquid phase, also are contemplated. The "beads" can include additional components, such as magnetic or paramagnetic particles (see, *e.g.*, Dynabeads (Dynal, Oslo, Norway)) for separation

using magnets, as long as the additional components do not interfere with the methods and analyses herein.

As used herein, a matrix or support particles refers to matrix materials that are in the form of discrete particles. The particles have any shape and dimensions, but typically have at least one dimension that is 100 mm or less, 50 mm or less, 10 mm or less, 1 mm or less, 100 $\mu$m or less, 50 $\mu$m or less and typically have a size that is 100 $mm^3$ or less, 50 $mm^3$ or less, 10 $mm^3$ or less, and 1 $mm^3$ or less, 100 $\mu m^3$ or less and can be order of cubic microns. Such particles are collectively called "beads."

As used herein, the abbreviations for any protective groups, amino acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, *Biochem.*, 11:942-944, 1972).

**B. Directed Evolution**

To date, there have been three general approaches described for protein directed evolution based on mutagenesis.

**1) Pure Random Mutagenesis**

Random mutagenesis methodology requires that the amino acids in the starting protein sequence are replaced by all (or a group) of the 20 amino acids. Either single or multiple replacements at different amino acid positions are generated on the same molecule, at the same time. The random mutagenesis method relies on a direct search for fitness improvement based on random amino acid replacement and sequence changes at multiple amino acid positions. In this approach neither the amino acid position (first dimension) nor the amino acid type (second dimension) are restricted; and everything possible is generated and tested. Multiple replacements can randomly happen at the same time on the same molecule. For example, random mutagenesis methods are widely used to develop antibodies with higher affinity for its ligand, by the generation of

random-sequence libraries of antibody molecules, followed by expression and screening using filamentous phages.

**2) Restricted Random Mutagenesis**

Restricted random mutagenesis methods introduce either all of the 20 amino acids or DNA-biased residues. The bias is based on the sequence of the DNA and not on that of the protein, in a stochastic or semi-stochastic manner, respectively, within restricted or predefined regions of the protein, known in advance to be involved in the biological activity being "evolved." This method relies on a direct search for fitness improvement based on random amino acid replacement and sequence changes at either restricted or multiple amino acid positions. In this approach the scanning can be restricted to selected amino acid positions and/or amino acid types, while material changes continue to be random in position and type. For example, the amino acid position can be restricted by prior selection of the target region to be mutated (selection of target region is based upon prior knowledge on protein structure/function); while the amino acid type is not primarily restricted as replacing amino acids are stochastically or at most "semi-stochastically" chosen. As an example, this method is used to optimize known binding sites on proteins, including hormone-receptor systems and antibody-epitope systems.

**3) Non-restricted Rational mutagenesis**

Rational mutagenesis is a two-step process and is described in co-pending U.S. application Serial No. 10/022,249. Briefly, the first step requires amino acid scanning where all and each of the amino acids in the starting protein sequence are replaced by a third amino acid of reference (e.g., alanine). Only a single amino acid is replaced on each protein molecule at a time. A collection of protein molecules having a single amino acid replacement is generated such that molecules differ from each other by the amino acid position at which the replacement has taken place. Mutant DNA molecules are designed, generated by mutagenesis and cloned individually, such as in addressable arrays, such that they are

physically separated from each other and such that each one is the single product of an independent mutagenesis reaction. Mutant protein molecules derived from the collection of mutant nucleic acid molecules also are physically separated from each other, such as by formatting in addressable arrays. Activity assessment on each protein molecule allows for the identification of those amino acid positions that result in a drop in activity when replaced, thus indicating the involvement of that particular amino acid position in the protein's biological activity and/or conformation that leads to fitness of the particular feature being evolved. Those amino acid positions are referred to as HITs. At the second step, a new collection of molecules is generated such that each molecule differs from each of the others by the amino acid present at the individual HIT positions identified in step 1. All 20 amino acids (19 remaining) are introduced at each of the HIT positions identified in step 1; while each individual molecule contains, in principle, one and only one amino acid replacement. Mutant DNA molecules are designed, generated by mutagenesis and cloned individually, such as in in addressable arrays, such that they are physically separated from each other and such that each one is the single product of an independent mutagenesis reaction. Mutant protein molecules derived from the collection of mutant DNA molecules also are physically separated from each other, such as by formatting in addressable arrays. Activity assessment then is individually performed on each individual mutant molecule. The newly generated mutants that lead to a desired alteration (such as an improvement) in a protein activity are referred to as LEADs. This method permits an indirect search for activity alteration, such as improvement, based on one rational amino acid replacement and sequence change at a single amino acid position at a time, in search of a new, unpredicted amino acid sequence at some unpredicted regions along a protein to produce a protein that exhibits a desired activity or altered activity, such as better performance than the starting protein.

In this approach, neither the amino acid position nor the replacing amino acid type are restricted. Full length protein scanning is performed during the first step to identify HIT positions, and then all 20 amino acids are tested at each of the HIT positions, to identify LEAD sequences; while, as a starting point, only one amino acid at a time is replaced on each molecule. The selection of the target region (HITs and surrounding amino acids) for the second step is based upon experimental data on activity obtained in the first step. Thus, no prior knowledge of protein structure and/or function is necessary. Using this approach, LEAD sequences have been found on proteins that are located at regions of the protein not previously known to be involved in the particular biological activity being optimized; thus emphasizing the power of this approach to discover unpredictable regions (HITs) as targets for fitness improvement.

**C. 2-Dimensional Rational Scanning**

The 2-Dimensional rational scanning (or "2-dimensional scanning") methods for protein rational evolution provided herein (see, also copending U.S. application Serial No. Attorney docket no. 923, filed the same day herewith, based on U.S. provisional application Serial Nos. 60/457,063 and 60/410,258) are based on scanning over two dimensions. The first dimension scanned is amino acid position along the protein sequence to identify is-HIT target positions, and the second dimension is the amino acid type selected for replacing a particular is-HIT amino acid position. An advantage of the 2-dimensional scanning methods provided herein is that at least one, and typically both, of the amino acid position scan and/or the replacing amino acid scan can be restricted such that fewer than all amino acids on the protein-backbone are selected for amino acid replacement; and/or fewer than all of the remaining 19 amino acids available to replace an original, such as native, amino acid are selected for replacement.

In particular embodiments, based on *i*) the particular protein properties to be evolved, *ii*) the protein's amino acid sequence, and *iii*) the

known properties of the individual amino acids, a number of target positions along the protein sequence are selected, *in silico*, as "is-HIT target positions." This number of is-HIT target positions is as large as possible such that all reasonably possible target positions for the particular feature being evolved are included. In particular, embodiments where a restricted number of is-HIT target positions are selected for replacement, the amino acids selected to replace the is-HIT target positions on the particular protein being optimized can be either all of the remaining 19 amino acids or, more frequently, a more restricted group comprising selected amino acids that are contemplated to have the desired effect on protein activity. In another embodiment, so long as a restricted number of replacing amino acids are used, all of the amino acid positions along the protein backbone can be selected as is-HIT target positions for amino acid replacement. Mutagenesis then is performed by the replacement of single amino acid residues at specific is-HIT target positions on the protein backbone (e.g., "one-by-one," such as in addressable arrays), such that each individual mutant generated is the single product of each single mutagenesis reaction. Mutant DNA molecules are designed, generated by mutagenesis and cloned individually, such as in addressable arrays, such that they are physically separated from each other and that each one is the single product of an independent mutagenesis reaction. Mutant protein molecules derived from the collection of mutant DNA molecules also are physically separated from each other, such as by formatting in in addressable arrays. Thus, a plurality of mutant protein molecules are produced. Each mutant protein contains a single amino acid replacement at only one of the is-HIT target positions. Activity assessment is then individually performed on each individual protein mutant molecule, following protein expression and measurement of the appropriate activity. An example of practice of this this method is shown in the Example in which mutant IFN$\alpha$ molecules and IFN$\beta$ moleucles are produced.

The newly generated proteins that lead to altered, typically improvement, in a target protein activity are referred to as LEADs. This method relies on an indirect search for protein improvement for a particular activity, such as increased resistance to proteolysis, based on a rational amino acid replacement and sequence change at single or, in another embodiment, a limited number of amino acid positions at a time. As a result, optimized proteins that have new amino acid sequences at some regions along the protein that perform better (at a particular target activity or other property) than the starting protein are identified and isolated.

**1) Identifying *in-silico* HITs**

Provided herein is a method for directed evolution that includes identifying and selecting (using *in silico* analysis) specific amino acids and amino acid positions (referred to herein as is-HITs) along the protein sequence that are contemplated to be directly or indirectly involved in the feature being evolved. As noted, the 2-dimensional scanning methods provided include the following two-steps. The first step is an *in silico* search of a target protein's amino acid sequence to identify all possible amino acid positions that potentially can be targets for the activity being evolved. This is effected, for example, by assessing the effect of amino acid residues on the property(ies) to be altered on the protein, using any known standard software. The particulars of the *in silico* analysis is a function of the property to be modified. For example, in the example herein, a property that is altered resistance of the protein to proteolysis.. To determine aminoacid residues that are potential targets as is-HITs, in this example, all possible target residues for proteases were first identified. The 3-dimensional structure of the protein was then considered in order to identify surface residues. Comparison of exposed residues with proteolytically cleavable residues yields residues that are targets for change.

Once identified, these amino acid positions or target sequences are referred to as "is-HITs" (*in silico* HITs). *In silico* HITs are defined as those amino acid positions (or target positions) that potentially are involved in the "evolving" feature, such as increased resistance to proteolysis. In one embodiment, the discrimination of the is-HITs among all the amino acid positions in a protein sequence is made based on *i*) the amino acid type at each position in addition to, whenever available but not necessarily, *ii*) the information on the protein secondary or tertiary structure. *In silico* HITs constitute a collection of mutant molecules such that all possible amino acids, amino acid positions or target sequences potentially involved in the evolving feature are represented. No strong theoretical discrimination among amino acids or amino acid positions is made at this stage.

*In silico* HIT positions are spread over the full length of the protein sequence. In one embodiment, only a single is-HIT amino acid at a time is replaced on the target protein. In another embodiment, a limited number of is-HIT amino acids are replaced at the same time on the same target protein molecule. The selection of target regions (is-HITs and surrounding amino acids) for the second step is based upon rational assumptions and predictions. No prior knowledge of protein structure/function is necessary. Hence, the 2-dimensional scanning methodology provided herein does not require any previous knowledge of the 3-dimensional conformational structure of the protein.

Any protein known or otherwise available to those of skill in the art is suitable for modification using the directed evolution methods provided herein, including cytokines (e.g., IFN$\alpha$-2b) or any other proteins that have previously been mutated or optimized.

A variety of parameters can be analyzed to determine whether or not a particular amino acid on a protein might be involved in the evolving feature. For example, the information provided by crystal structures of proteins can be rationally exploited in order to perform a computer-

assisted (*in silico*) analysis towards the prediction of variants with desired features. In a particular embodiment, a limited number of initial premises (typically no more than 2) are used to determine the *in silico* HITs. In other embodiments, the number of premises used to determine the in silico hits can range from 1 to 10 premises, including no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, but are typically no more than 2 premises. It is important to the methods provided herein that the number of initial premises be kept to a minimum, so as to maintain the number of potential is-HITs at a maximum (here is where the methods provided are not limited by too much prediction based on theoretical assumptions). When two premises are employed, the first condition is typically the amino acid type itself, which is directly linked to the nature of the evolving feature. For example, if the goal were to change the optimum pH for an enzyme, then the replacing amino acids selected at this step for the replacement of the original sequence would be only those with a certain pKa value. The second premise is typically related to the specific position of those amino acids along the protein structure. For example, some amino acids might be discarded if they are not expected to be exposed enough to the solvent, even when they might have appropriate pKa values.

During the first step of identification of is-HITs according to the methods provided herein, each individual amino acid along the protein sequence is considered individually to assess whether it is a candidate for is-HIT. This search is done one-by-one and the decision on whether the amino acid is considered to be a candidate for a is-HIT is based on (1) the amino acid type itself; (2) the position on the amino acid sequence and protein structure if known; and (3) the predicted interaction between that amino acid and its neighbors in sequence and space.

Using the 3D-scanning methods provided herein, once one protein within a family of proteins (e.g., IFN$\alpha$-2b within the cytokine family) is optimized using the methods provided herein for generating LEAD

mutants, is-HITs can be identified on other or all proteins within a particular family by identifying the corresponding amino acid positions therein using structural homology analysis (based upon comparisons of the 3-D structures of the family members with original protein to identify corresponding residues for replacement) as described hereinafter. The is-HITs on family identified in this manner then can be subjected to the next step of identifying replacing amino acids and further assayed to obtain LEADs or super-LEADs as described herein.

**2) Identifying Replacing Amino Acids**

Once the is-HITs target positions are selected, the next step is identifying those amino acids that will replace the original, such as native, amino acid at each is-HIT position to alter the activity level for the particular feature being evolved. The set of replacing amino acids to be used to replace the original, such as native, amino acid at each is-HIT position can be different and specific for the particular is-HIT position. The choice of the replacing amino acids takes into account the need to preserve the physicochemical properties such as hydrophobicity, charge and polarity, of essential (e.g., catalytic, binding, etc.) residues. The number of replacing amino acids, of the remaining 19 non-native (or non-original) amino acids, that can be used to replace a particular is-HIT target position ranges from 1 up to about 19, from 1 up to about 15, from 1 up to about 10, from 1 up to about 9, from 1 up to about 8, from 1 up to about 7, from 1 up to about 6, from 1 up to about 5, from 1 up to about 4, from 1 up to about 3, or from 1 to 2 amino acid replacements.

Numerous methods of selecting replacing amino acids (also referred to herein as "replacement amino acids") are well known in the art. Protein chemists determined that certain amino acid substitutions commonly occur in related proteins from different species. As the protein still functions with these substitutions, the substituted amino acids are compatible with protein structure and function. Often, these substitutions

are to a chemically similar amino acid, but other types of changes, although relatively rare, can also occur.

Knowing the types of changes that are most and least common in a large number of proteins can assist with predicting alignments and amino acid substitutions for any set of protein sequences. Amino acid substitution matrices are used for this purpose.

In amino acid substitution matrices, amino acids are listed across the top of a matrix and down the side, and each matrix position is filled with a score that reflects how often one amino acid would have been paired with the other in an alignment of related protein sequences. The probability of changing amino acid A into amino acid B is assumed to be identical to the reverse probability of changing B into A. This assumption is made because, for any two sequences, the ancestor amino acid in the phylogenetic tree is usually not known. Additionally, the likelihood of replacement should depend on the product of the frequency of occurrence of the two amino acids and on their chemical and physical similarities. A prediction of this model is that amino acid frequencies will not change over evolutionary time (Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978). Below are several exemplary amino acid substitution matrices, including, but not limited to block substitution matrix (BLOSUM), Jones, Gonnet, Fitch, Feng, McLachlan, Grantham, Miyata, Rao, Risler, Johnson and percent accepted mutation (PAM). Any such method known to those of skill in the art can be employed.

**(a) Percent accepted mutation (PAM)**

Dayhoff and coworkers developed a model of protein evolution that resulted in the development of a set of widely used replacement matrices (Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978) termed percent accepted mutation matrices (PAM). In deriving these matrices, each change in the current amino acid at a particular site is assumed to be independent of previous mutational events at that site. Thus, the probability of change of any amino acid A to amino acid B is

the same, regardless of the previous changes at that site and also regardless of the position of amino acid A in a protein sequence.

In the Dayhoff approach, replacement rates are derived from alignments of protein sequences that are at least 85% identical; this constraint ensures that the likelihood of a particular mutation being the result of a set of successive mutations is low. Because these changes are observed in closely related proteins, they represent amino acid substitutions that do not significantly change the function of the protein. Hence, they are called "accepted mutations," as defined as amino acid changes that are accepted by natural selection.

**(i) PAM Analysis**

In particular embodiments of the methods provided herein, "Percent Accepted Mutation" (PAM; Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978 FIG2) PAM values are used to select an appropriate group of replacement amino acids. PAM matrices were originally developed to produce alignments between protein sequences based evolutionary distances. Because, in a family of proteins or homologous (related) sequences, identical or similar amino acids (85% similarity) are shared, conservative substitutions for, or allowed point mutations of the corresponding amino acid residues can be determined throughout an aligned reference sequence. Conservative substitutions of a residue in a reference sequence are those substitutions that are physically and functionally similar to the corresponding reference residues, e.g., that have a similar size, shape, electric charge, chemical properties, including the ability to form bonds such as covalent and hydrogen bonds. Particularly suitable conservative amino acid substitutions are those that show the highest scores and fulfill the PAM matrix criteria in the form of "accepted point mutations." For example, by comparing a family of scoring matrices, Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978, found a consistently higher

score significance when using PAM250 matrix to analyze a variety of proteins, known to be distantly related.

**(ii) PAM 250**

In a particular embodiment, the PAM250 matrix set forth in FIG2 is used for determining the replacing amino acids based on similarity criteria. The PAM250 matrix uses data obtained directly from natural evolution to facilitate the selection of replacing amino acids for the is-HITs to generate conservative mutations without much affecting the overall protein function. By using the PAM250 matrix, candidate replacing amino acids are identified from related proteins from different organisms.

**(b) Jones and Gonnet**

This method (see, *e.g.*, Jones *et al.*, *Comput. Appl. Biosci.*, 8:275-282, 1992 and Gonnet *et al.*, *Science*, 256:1433-1445, 1992) uses much of the same methodology as Dayhoff (see below), but with modern databases. The matrix of Jones *et al.*, is extracted from Release 15.0 of the SWISS-PROT protein sequence database. Point mutations totaling 59,160 from 16,130 protein sequences were used to calculate a PAM250 (see below) matrix.

The matrix published by Gonnet *et al.*, *Science*, 256:1433-1445, 1992, was built from a sequence database of 8,344,353 amino acid residues. Each sequence was compared against the entire database, such that $1.7 \times 10^6$ subsequent matches resulted for the significant alignments. These matches were then used to generate a matrix with a PAM distance of 250.

**(c) Fitch and Feng**

Fitch, *J. Mol. Evol.*, 16(1):9-16, 1966, used an exchange matrix that contained for each pair (A, B) of amino acid types the minimum number of nucleotides that must be changed to encode amino acid A instead of amino acid B. Feng *et al.*, *J. Mol. Evol.*, 21:112-125, 1985, used an enhanced version of Fitch, *J. Mol. Evol.*, 16(1):9-16, 1966, to build a Structure-Genetic matrix. In addition to considering the minimum

number of base changes required to encode amino acid B instead of A, this method also considers the structural similarity of the amino acids.

**(d) McLachlan, Grantham and Miyata**

McLachlan, *J. Mol. Biol.*, 61:409-424 1971, used 16 protein families, each with 2 to 14 members. The 89 sequences were aligned and the pairwise exchange frequency, observed in 9280 substitutions, was used to generate an exchange matrix with values varying from 0 to 9.

Grantham, *Science*, 185:862-864, 1974, considers composition, polarity and molecular volume of amino acid side-chains, properties that were highly correlated to the relative substitution frequencies tabulated by McLachlan, *J. Mol. Biol.*, 61:409-424, 1971, to build the matrix.

Miyata, *J. Mol. Evol.*, 12:219-236, 1979, uses the volume and polarity values of amino acids published by Grantham, *Science*, 185:862-864, 1974. For every amino acid type pair, the difference for both properties was calculated and divided by the standard deviation of all the differences. The square root of the sum of both values is then used in the matrix.

**(e) Rao**

Rao, *J. Pept. Protein Res.*, 29:276-281, 1987, employs five amino acid properties to create a matrix; namely, alpha-helical, beta-strand and reverse-turn propensities as well as polarity and hydrophobicity. The standardized properties were summed and the matrix rescaled to the same average as that for PAM (Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978).

**(f) Risler**

Risler *et al.*, *J. Mol. Biol.*, 204:1019-1029, 1988, aligned 32 three-dimensional structures from 11 protein families by rigid-body superposition of the backbone topology. Only substitutions were considered where at least three adjacent and equivalent main-chain $C^{\alpha}$ atom pairs in the compared structures were each not more than 1.2 Å

apart. A total of 2860 substitutions were considered and used to build a matrix based on $\chi^2$ distance calculations.

**(g) Johnson**

Johnson *et al.*, *J. Mol. Biol.*, 233:716-738, 1993, derived their matrix from the tertiary structural alignment of 65 families in a database of 235 structures created with the method of Sali *et al.*, *J. Mol. Biol.*, 212:403-428, 1990. Their examination of the substitutions was based on the expected and observed ratios of occurrences and the final matrix values were taken as $\log_{10}$ of the ratios.

**(h) Block Substitution Matrix (BLOSUM)**

One empirical approach (Henikoff *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:10915-10919, 1992) uses local, ungapped alignments of distantly related sequences to derive the blocks amino acid substitution matrix (BLOSUM) series of matrices. The matrix values are based on the observed amino acid substitutions in a larger set of about 2000 conserved amino acid patterns, termed blocks. These blocks act as signatures of families of related proteins. Matrices of this series are identified by a number after the matrix (e.g., BLOSUM50), which refers to the minimum percentage identity of the blocks of multiple aligned amino acids used to construct the matrix. It is noteworthy that these matrices are directly calculated without extrapolations, and are analogous to transition probability matrices P(T) for different values of T, estimated without reference to any rate matrix Q.

The outcome of these two steps set forth above, which is performed *in silico* is that: (1) the amino acid positions that will be the target for mutagenesis are identified; these positions are referred to as is-HITs; (2) the replacing amino acids for the original, such as native, amino acids at the is-HITs are identified, to provide a collection of candidate LEAD mutant molecules that are expected to perform different from the native one. These are are assayed for a desired optimized (or improved or altered) biological activity.

### 3) Physical Construction of Mutant Proteins and Biological Assays

Once is-HITs are selected as set forth above, replacing amino acids are introduced. Mutant proteins typically are prepared using recombinant DNA methods and assessed in appropriate biological assays for the particular biological activity (feature) optimized (see, *e.g.*, Example 1). An exemplary method of preparing the mutant proteins is by mutagenesis of the original, such as native, gene using methods well known in the art. Mutant molecules are generated one-by-one, such as in addressable arrays, such that each individual mutant generated is the single product of each single and independent mutagenesis reaction. Individual mutagenesis reactions are conducted separately, such as in addressable arrays where they are physically separated from each other. Once a population of sets of nucleic acid molecules encoding the respective mutant proteins is prepared, each is separately introduced one-by-one into appropriate cells for the production of the corresponding mutant proteins. This can also be performed, for example, in addressable arrays where each set of nucleic acid molecules encoding a respective mutant protein is introduced into cells confined to a discrete location, such as in a well of a multi-well microtiter plate. Each individual mutant protein is individually phenotypically characterized and performance is quantitatively assessed using assays appropriate for the feature being optimized (i.e., feature being evolved). Again, this step can be performed in addressable arrays. Those mutants displaying a desired increased or decreased performance compared to the original, such as native molecules are identified and designated LEADs. From the beginning of the process of generating the mutant DNA molecules up through the readout and analysis of the performance results, each candidate LEAD mutant is generated, produced and analyzed individually, such as from its own address in an addressable array. The process is amenable to automation.

## D. 2-Dimensional Scanning of Proteins for Incr ased Resistanc to Proteolysis

The methods of 2-dimensional scanning permit preparation of proteins modified for a selected trait, activity or other phenotype. Among modifications of interest for therapeutic proteins are those that increase protection against protease digestion while maintaining the requisite biological activity. Such changes are useful for producing longer-lasting therapeutic proteins.

The delivery of stable peptide and protein drugs to patients is a major challenge for the pharmaceutical industry. These types of drugs in the human body are constantly eliminated or taken out of circulation by different physiological processes including internalization, glomerular filtration and proteolysis. The latter is often the limiting process affecting the half-life of proteins used as therapeutic agents in per-oral administration and either intravenous or intramuscular injections.

The 2-dimensional scanning process for protein evolution is used to effectively improve protein resistance to proteases and thus increase protein half-life *in vitro* and, ultimately *in vivo*. As noted, the methods provided herein for designing and generating highly stable, longer lasting proteins, or proteins having a longer half-life include: *i*) identifying some or all possible target sites on the protein sequence that are susceptible to digestion by one or more specific proteases (these sites are referred to herein as is-HITs); *ii*) identifying appropriate replacing amino acids, specific for each is-HIT, such that upon replacement of one or more of the original, such as native, amino acids at that specific is-HIT, they can be expected to increase the is-HIT's resistance to digestion by protease while at the same time, maintaining or improving the requisite biological activity of the protein (these proteins with replaced amino acids are the "candidate LEADs"); *iii*) systematically introducing the specific replacing amino acids (candidate LEADs) at every specific is-HIT target position to generate a collection containing the corresponding mutant candidate lead

molecules. Mutants are generated, produced and phenotypically characterized one-by-one, such as in addressable arrays, such that each mutant molecule contains initially an amino acid replacement at only one is-HIT site.

In particular embodiments, such as in subsequent rounds, mutant molecules also can be generated that contain one or more amino acids at one or more is-HIT sites that have been replaced by candidate LEAD amino acids. Those mutant proteins carrying one or more mutations at one or more is-HITs, and that display improved protease resistance are called LEADs (one mutation at one is-HIT) and super-LEADs (mutations at more than one is-HIT).

The first step of the process takes into consideration existing knowledge from different domains:

(1) About the galenic and the delivery environment (tissue, organ or corporal fluid) of the particular therapeutic protein in order to establish a list of proteases more likely to be found in that environment. For example, a therapeutic protein in per-oral application is likely to encounter typical proteases of the luminal gastrointestinal tract. In contrast, if this protein were injected in the blood circulation, serum proteases would be implicated in the proteolysis. Based on the specific list of proteases involved, the complete list of all amino acid sequences that potentially could be targeted by the proteases in the list is determined.

(2) Since protease mixtures in the body are quite complex in composition, almost all the residues in any target protein potentially are targeted for proteolysis (FIG6A). Nevertheless, proteins form specific tri-dimensional structures where residues are more or less exposed to the environment and protease action. It can be assumed that those residues constituting the core of a protein are inaccessible to proteases, while those more 'exposed' to the environment are better targets for proteases. The probability for every specific amino acid to be 'exposed' and then to

be accessible to proteases can be taken into account to reduce the number of is-HIT. Consequently, the methods herein consider the analysis with respect to solvent "exposure" or "accessibility" for each individual amino acid in the protein sequence. Solvent accessibility of residues can alternatively be estimated, regardless of any previous knowledge of specific protein structural data, by using an algorithm derived from empirical amino acid probabilities of accessibility, which is expressed in the following equation (Boger *et al.*, *Reports of the Sixth International Congress in Immunology*, p. 250, 1986):

$$A(i) = [\prod_{j=1}^{6} \delta_{i+4-j}] * [0.62]^{-6}.$$

Briefly, these are fractional probabilities ($\delta_{(i)}$) determined for an amino acid (i) found on the surface of a protein, which are based upon structural data from a set of several proteins. It is thus possible to calculate the solvent accessibility (A) of an amino acid (A(i)) at sequence position (i-2 to i+3, onto a sliding window of length equal to 6) that is within an average surface accessible to solvent of >20 square angstroms ($Å^2$).

The protease accessible target amino acids along the protein sequence, i.e., the amino acids to be replaced, are thus identified and are referred to herein as *in silico* HITs (is-HITs).

Amino acids at the is-HITs then are replaced by residues that render the sequence less vulnerable (by a factor, for example, of 1%, 10%, 20%, 30%, 40%, 50%, . . . 100% depending upon the protein) or invulnerable (substantially no detectable digestion within a set time period) to protease digestion, while at the same time maintain a biological activity or activities of interest of the protein. The choice of the replacing amino acids is complicated by (1) the broad target specificity of certain proteases and (2) the need to preserve the physicochemical properties such as hydrophobicity, charge and polarity, of essential (e.g., catalytic,

binding and/or other activities depending upon the protein) residues. For use in the methods herein, the "Percent Accepted Mutation" values (PAM values; see, Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978), FIG2) can be used as desribed herein. PAM values, originally developed to produce alignments between protein sequences, are available in the form of probability matrices, which reflect an evolutionary distance. Since, in a family of proteins or homologous (related) sequences, identical or similar amino acids (85% similarity) are shared, conservative substitutions for, or "allowed point mutations" of the corresponding amino acid residues can be determined throughout an aligned reference sequence. As noted, conservative substitutions of a residue in a reference sequence are those substitutions that are physically and functionally similar to the corresponding reference residues e.g., that have a similar size, shape, electric charge, chemical properties,2 including the ability to form bonds such as covalent and hydrogen bonds. For example, conservative substitutions can be those that exhibit the highest scores and fulfill the PAM matrix criteria in the form of "accepted point mutations."

By comparing a family of scoring matrices, Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, 5(3):345-352, 1978), found consistently higher score significance when using PAM250 matrix to analyze a variety of proteins, known to be distantly related. For methods herein, the PAM250 matrix was selected for use. The PAM250 matrix is used, by learning directly from natural evolution, to find replacing amino acids for the is-HITs to generate conservative mutations without affecting the protein function. By using PAM250, candidate replacing amino acids are identified from related proteins from different organisms.

An exemplary class of proteins that can be optimized according to the methods provided herein are the cytokines. For example, 2D-scanning methods provided herein can be used to modify the following cytokines to increase their stability as assessed by an increased

resistance to proteolysis resulting in an increased protein half-life in the bloodstream or any other desired biological activity of the selected protein. Exemplary cytokines, include, but are not limited to: interleukin-10 (IL-10; SEQ ID NO: 200), interferon beta (IFN$\beta$; SEQ ID NO: 196), interferon alpha-2a (IFN$\alpha$-2a; SEQ ID NO: 182), interferon alpha-2b (IFN$\alpha$-2b; SEQ ID NO: 1), and interferon gamma (IFN-$\gamma$; SEQ ID NO: 199), granulocyte colony stimulating factor (G-CSF; SEQ ID NO: 210), leukemia inhibitory factor (LIF; SEQ ID NO: 213), growth hormone (hGH; SEQ ID NO: 216), ciliary neurotrophic factor (CNTF; SEQ ID NO: 212), leptin (SEQ ID NO: 211), oncostatin M (SEQ ID NO: 214), interleukin-6 (IL-6; SEQ ID NO: 217), interleukin-12 (IL-12; SEQ ID NO: 215), erythropoietin (EPO; SEQ ID NO: 201), granulocyte-macrophage colony stimulating factor (GM-CSF; SEQ ID NO: 202), interleukin-2 (IL-2; SEQ ID NO: 204), interleukin-3 (IL-3; SEQ ID NO: 205), interleukin-4 (IL-4; SEQ ID NO: 207), interleukin-5 (IL-5; SEQ ID NO: 208), interleukin-13 (IL-13; SEQ ID NO: 209), Flt3 ligand (SEQ ID NO: 203) and stem cell factor (SCF; SEQ ID NO: 206).

Accordingly, provided herein are modified cytokines that exhibit increased resistance to proteolysis compared to the unmodified cytokine. The modified cytokines can be selected from among a member of the interferons/interleukin-10 protein family, a member of the long-chain cytokine family; and a member of the short-chain cytokine family. In particular embodiments, the modified cytokines provided herein are selected from among: interleukin-10 (IL-10), interferon beta (IFN$\beta$), interferon alpha-2a (IFN$\alpha$-2a), interferon alpha-2b (IFN$\alpha$-2b), and interferon gamma (IFN-$\gamma$), granulocyte colony stimulating factor (G-CSF), leukemia inhibitory factor (LIF), human growth hormone (hGH), ciliary neurotrophic factor (CNTF), leptin, oncostatin M, interleukin-6 (IL-6) and interleukin-12 (IL-12), erythropoietin (EPO), granulocyte-macrophage colony stimulating factor (GM-CSF), interleukin-2 (IL-2), interleukin-3 (IL-3), interleukin-4 (IL-4), interleukin-5 (IL-5), interleukin-13 (IL-13), Flt3 ligand and stem cell

factor (SCF). In one embodiment, the modified cytokine is an interferon, including modified interferon $\alpha$-2b (IFN$\alpha$-2b).

**E. Rational Evolution of IFN$\alpha$-2b For Increased Resistance to Proteolysis**

IFN$\alpha$-2b is used for a variety of applications. Typically it is used for treatment of type B and C chronic hepatitis. Additional indications include, but are not limited to, melanomas, herpes infections, Kaposi sarcomas and some leukemia and lymphoma cases. Patients receiving interferon are subject to frequent repeat applications of the drug. Since such frequent injections generate uncomfortable physiological as well as undesirable psychological reactions in patients, increasing the half-life of interferons and thus decreasing the necessary frequency of interferon injections, would be extremely useful to the medical community. For example, after injection of native human IFN$\alpha$-2b injection in mice, as a model system, its presence can be detected in the serum between 3 and 10 hours with a half-life of only around 4 hours. The IFN$\alpha$-2b completely disappears to undetectable levels by 18-24 hours after injection. Provided herein are mutant variants of the IFN$\alpha$-2b protein that display altered properties including: (a) highly improved stability as assessed by resistance to proteases in vitro and by pharmacokinetics studies in mice; and (b) at least comparable biological activity as assessed by antiviral and antiproliferative action compared to both the unmodified and wild type native IFN$\alpha$-2b protein and to at least one pegylated derivative of the wild type native IFN$\alpha$. As a result, the IFN$\alpha$-2b mutant proteins provided herein confer a higher half-life and at least comparable antiviral and antiproliferation activity (sufficient for a therapeutic effect) with respect to the native sequence and to the pegylated derivatives molecules currently being used for the clinical treatment of hepatitis C infection. See Figures 6(A)-6(N), 6(T) and 6(U). Thus, the optimized IFN$\alpha$-2b protein mutants that possess increased resistance to proteolysis and/or glomerular filtration provided herein result in a decrease in the frequency of injections

needed to maintain a sufficient drug level in serum, leading to *i*) higher comfort and acceptance by patients, *ii*) lower doses necessary to achieve comparable biological effects, and *iii*) as a consequence of (*ii*), an attenuation of the (dose-dependent) secondary effects observed in humans.

In particular embodiments, the half-life of the IFN$\alpha$-2b and IFN$\alpha$-2a mutants provided herein is increased by an amount selected from at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500% or more, when compared to the half-life of native human IFN$\alpha$-2b and IFN$\alpha$-2a in either human blood, human serum or an in vitro mixture containing one or more proteases. In other embodiments, the half-life of the IFN$\alpha$-2b and IFN$\alpha$-2a mutants provided herein is increased by an amount selected from at least 6 times, 7 times, 8 times, 9 times, 10 times, 20 times, 30 times, 40 times, 50 times, 60 times, 70 times, 80 times, 90 times, 100 times, 200 times, 300 times, 400 times, 500 times, 600 times, 700 times, 800 times, 900 times, 1000 times, or more, when compared to the half-life of native human IFN$\alpha$-2b and IFN$\alpha$-2a in either human blood, human serum or an in vitro mixture containing one or more proteases.

Two methodologies were used herein to increase the stability of IFN$\alpha$-2b by amino acid replacement: *i*) amino acid replacement that leads to higher resistance to proteases by direct destruction of the protease target residue or sequence, while either maintaining or improving the requisite biological activity (e.g., antiviral activity, antiproliferation activity), and/or *ii*) amino acid replacement that leads to a different pattern of *N*-glycosylation, thus decreasing both glomerular filtration and sensitivity to proteases, while either improving or maintaining the requisite biological activity (e.g., antiviral activity, antiproliferation activity).

The 2D-scanning methods provided herein were used to identify the amino acid changes on IFN$\alpha$-2b that lead to an increase in stability when challenged either with proteases, human blood lysate or human serum. Increasing protein stability to proteases, human blood lysate or human serum, and/or increasing the molecular size is contemplated herein to provide a longer *in vivo* half-life for the particular protein molecules, and thus to a reduction in the frequency of necessary injections into patients. The biological activities that were measured for the IFN$\alpha$-2b molecules are *i*) their capacity to inhibit virus replication when added to permissive cells previously infected with the appropriate virus, and *ii*) their capacity to stimulate cell proliferation when added to the appropriate cells. Prior to the measurement of biological activity, IFN$\alpha$-2b molecules were challenged with proteases, human blood lysate or human serum during different incubation times. The biological activity measured, corresponds then to the residual biological activity following exposure to the protease-containing mixtures.

As set forth above, provided herein are methods for the development of IFN$\alpha$-2b and IFN$\alpha$-2a molecules that, while maintaining the requisite biological activity intact, have been rendered less susceptible to digestion by blood proteases and therefore display a longer half-life in blood circulation. In this particular example, the method used included the following specific steps as set forth in Example 2:

1) Identifying some or all possible target sites on the protein sequence that are susceptible to digestion by one or more specific proteases (these sites are the is-HITs) and

2) Identifying appropriate replacing amino acids, specific for each is-HIT, such that if used to replace one or more of the original amino acids at that specific is-HIT, they can be expected to increase the is-HIT's resistance to digestion by protease while at the same time, keeping the biological activity of the protein

unchanged (these replacing amino acids are the "candidate LEADs").

As set forth in Example 2, the 3-dimensional structure of IFN$\alpha$-2b obtained from the NMR structure of IFN$\alpha$-2a (PDB code 1ITF) was used to select only those residues exposed to solvent from a list of residues along the IFN$\alpha$-2b and IFN$\alpha$-2a sequence which can be recognized as a substrate for different enzymes present in the serum. Residue 1 corresponds to the first residue of the mature peptide IFN$\alpha$-2b encoded by nucleotides 580-1074 of sequence accession No. J00207, SEQ ID NO:1. Using this approach, the following 42 amino acid target positions were identified as is-HITs on IFN$\alpha$-2b or IFN$\alpha$-2a, which numbering is that of the mature protein (SEQ ID NO:1 or SEQ ID NO:182, respectively): L3, P4, R12, R13, M16, R22, K23 or R23, F27, L30, K31, R33, E41, K49, E58, K70, E78, K83, Y89, E96, E107, P109, L110, M111, E113, L117, R120, K121, R125, L128, K131, E132, K133, K134, Y135, P137, M148, R149, E159, L161, R162, K164, and E165. Each of these positions was replaced by residues defined as compatible by the substitution matrix PAM250 while at the same time not generating any new substrates for proteases. For these 42 is-HITs, the residue substitutions determined by PAM250 analysis were as follows:

R to H, Q

E to H, Q

K to Q, T

L to V, I

M to I, V

P to A, S

Y to I, H.

### 1) Modified IFN$\alpha$-2b Proteins with Single Amino Acid Substitutions

Among the mutant proteins provided herein, are mutant IFN$\alpha$-2b proteins that have increased resistance proteolysis compared to the unmodified, typically wild-type, protein. The mutant IFN$\alpha$-2b proteins include those selected from among proteins containing more single amino acid replacements in SEQ ID NO:1, corresponding to: L by V at position 3; L by I at position 3; P by S at position 4; P by A at position 4; R by H at position 12; R by Q at position 12; R by H at position 13; R by Q at position 13; M by V at position 16; M by I at position 16; R by H at position 22; R by Q at position 22; R by H at position 23; R by Q at position 23; F by I at position 27; F by V at position 27; L by V at position 30; L by I at position 30; K by Q at position 31; K by T at position 31; R by H at position 33; R by Q at position 33; E by Q at position 41; E by H at position 41; K by Q at position 49; K by T at position 49; E by Q at position 58; E by H at position 58; K by Q at position 70; K by T at position 70; E by Q at position 78; E by H at position 78; K by Q at position 83; K by T at position 83; Y by H at position 89; Y by I at position 89; E by Q at position 96; E by H at position 96; E by Q at position 107; E by H at position 107; P by S at position 109; P by A at position 109; L by V at position 110; L by I at position 110; M by V at position 111; M by I at position 111; E by Q at position 113; E by H at position 113; L by V at position 117; L by I at position 117; R by H at position 120; R by Q at position 120; K by Q at position 121; K by T at position 121; R by H at position 125; R by Q at position 125; L by V at position 128; L by I at position 128; K by Q at position 131; K by T at position 131; E by Q at position 132; E by H at position 132; K by Q at position 133; K by T at position 133; K by Q at position 134; K by T at position 134; Y by H at position 135; Y by I at position 135; P by S at position 137; P by A at position 137; M by V at position 148; M by I at position 148; R by H at position 149; R by Q at

position 149; E by Q at position 159; E by H at position 159; L by V at position 161; L by I at position 161; R by H at position 162; R by Q at position 162; K by Q at position 164; K by T at position 164; E by Q at position 165; and E by H at position 165.

**2) LEAD Identification**

Next the specific replacing amino acids (candidate LEADs) are systematically introduced at every specific is-HIT position to generate a collection containing the corresponding mutant IFN$\alpha$-2b DNA molecules, as set forth in Example 2. The mutant DNA molecules were used to produce the corresponding mutant IFN$\alpha$-2b protein molecules by transformation or transfection into the appropriate cells. These protein mutants were assayed for (i) protection against proteolysis, (ii) antiviral and antiproliferation activity in vitro, (iii) pharmacokinetics in mice. Of particular interest are mutations that increase these activities of the IFN$\alpha$-2b mutant proteins compared to unmodified wild type IFN$\alpha$-2b protein and to pegylated derivates of the wild type protein. Based on the results obtained from these assays, each individual IFN$\alpha$-2b variant was assigned a specific activity. Those variant proteins displaying the highest stability and/or resistance to proteolysis were selected as LEADs. The candidate LEADs that possessed at least as much residual antiviral activity following protease treatment as the control, native IFN$\alpha$-2b, before protease treatment were selected as LEADs. The results are set forth in Table 2 of Example 2.

Using this method, the following mutants selected as LEADs are provided herein and correspond to the group of proteins containing one or more single amino acid replacements in SEQ ID NO:1, corresponding to: F by V at position 27; R by H at position 33; E by Q at position 41; E by H at position 41; E by Q at position 58; E by H at position 58; E by Q at position 78; E by H at position 78; Y by H at position 89; E by Q at position 107; E by H at position 107; P by A at position 109; L by V at position 110; M by V at position 111; E by Q at position 113; E by H at

position 113; L by V at position 117; L by I at position 117; K by Q at position 121; K by T at position 121; R by H at position 125; R by Q at position 125; K by Q at position 133; K by T at position 133; and E by Q at position 159; E by H at position 159. Among these are mutations that can have multiple effects. For exmple, among mutations described herein, are mutations that result in an increase of the IFN$\alpha$-2b activity as assessed by detecting the requisite biological activity.

Also provided are IFN$\alpha$-2b proteins that contain a plurality of mutations based on the LEADs (see, *e.g.*, Tables 6 and 7, EXAMPLE 5, which listscandidate LEADs and LEAD sites), are generated. These IFN$\alpha$-2b proteins have activity that is further optimized. Examples of such proteins are described in the EXAMPLES. Other combinations of mutations can be prepared and tested as described herein to identify other LEADs of interest, particularly those that have further increased IFN$\alpha$-2b antiviral activity or further increased resistance to proteolysis.

Also provided herein are modified IFN$\alpha$-2b or IFN$\alpha$-2a cytokines selected from among proteins comprising one or more single amino acid replacements in SEQ ID NOS:1 or 182, corresponding to the replacement of: N by D at position 45 (e.g., SEQ ID NO:978); D by G at position 94 (e.g, SEQ ID NO:979); G by R at position 102 (e.g., SEQ ID NO:980); A by G at position 139 (e.g., SEQ ID NO:981); or any combination thereof. These particular proteins have also been found herein to have increased resistance to proteolysis.

In another embodiment, IFN$\alpha$-2b and IFN$\alpha$-2a proteins that contain a plurality of mutations based on the LEADs (see Tables in the EXAMPLES, listing the candidate LEADs and LEAD sites), are produced to produce IFN$\alpha$-2b and IFN$\alpha$-2a proteins that have activity that is further optimized. Examples of such proteins are described herein. Other combinations of mutations can be prepared and tested as described herein to identify other LEADs of interest, particularly those that have further

increased IFN$\alpha$-2b and IFN$\alpha$-2a antiviral activity or further increased resistance to proteolysis.

**3) N-glycosylation Site Addition**

In additional embodiments, N-glycosylation sites can be added to increase resistance to proteolysis while maintaining or improving the requisite biological activity. Exemplary N-glycosylation mutants containing duo-amino acid replacements corresponding to the N-X-S or N-X-T consensus sequences are set forth in Example 3. Accordingly, provided herein are IFN$\alpha$-2b and IFN$\alpha$-2a mutant proteins having an increased resistance to proteolysis compared to unmodified IFN$\alpha$-2b and IFN$\alpha$-2a, selected from among proteins comprising one or more sets of duo-amino acid replacements in SEQ ID NO:1, corresponding to:

D by N at position 2 and P by S at position 4;

D by N at position 2 and P by T at position 4;

L by N at position 3 and Q by S at position 5;

L by N at position 3 and Q by T at position 5;

P by N at position 4 and T by S at position 6;

P by N at position 4 and T by T at position 6;

Q by N at position 5 and H by S at position 7;

Q by N at position 5 and H by T at position 7;

T by N at position 6 and S by S at position 8;

T by N at position 6 and S by T at position 8;

H by N at position 7 and L by S at position 9;

H by N at position 7 and L by T at position 9;

S by N at position 8 and G by S at position 10;

S by N at position 8 and G by T at position 10;

L by N at position 9 and S by S at position 11;

L by N at position 9 and S by T at position 11;

M by N at position 21 and K by S at position 23;

M by N at position 21 and K by T at position 23;

R by N at position 22 and I by S at position 24;

R by N at position 22 and I by T at position 24;
K or R by N at position 23 and S by S at position 25;
K or R by N at position 23 and S by T at position 25;
I by N at position 24 and L by S at position 26;
I by N at position 24 and L by T at position 26;
S by N at position 25 and F by S at position 27;
S by N at position 25 and F by T at position 27;
L by N at position 26 and S by S at position 28;
L by N at position 26 and S by T at position 28;
S by N at position 28 and L by S at position 30;
S by N at position 28 and L by T at position 30;
L by N at position 30 and D by S at position 32;
L by N at position 30 and D by T at position 32;
K by N at position 31 and R by S at position 33;
K by N at position 31 and R by T at position 33;
D by N at position 32 and H by S at position 34;
D by N at position 32 and H by T at position 34;
R by N at position 33 and D by S at position 35;
R by N at position 33 and D by T at position 35;
H by N at position 34 and F by S at position 36;
H by N at position 34 and F by T at position 36;
D by N at position 35 and G by S at position 37;
D by N at position 35 and G by T at position 37;
F by N at position 36 and F by S at position 38;
F by N at position 36 and F by T at position 38;
G by N at position 37 and P by S at position 39;
G by N at position 37 and P by T at position 39;
F by N at position 38 and Q by S at position 40;
F by N at position 38 and Q by T at position 40;
P by N at position 39 and E by S at position 41;
P by N at position 39 and E by T at position 41;

Q by N at position 40 and E by S at position 42;
Q by N at position 40 and E by T at position 42;
E by N at position 41 and F by S at position 43;
E by N at position 41 and F by T at position 43;
E by N at position 42 and G by S at position 44;
E by N at position 42 and G by T at position 44;
F by N at position 43 and N by S at position 45;
F by N at position 43 and N by T at position 45;
G by N at position 44 and Q by S at position 46;
G by N at position 44 and Q by T at position 46;
N by N at position 45 and F by S at position 47;
N by N at position 45 and F by T at position 47;
Q by N at position 46 and Q by S at position 48;
Q by N at position 46 and Q by T at position 48;
F by N at position 47 and K by S at position 49;
F by N at position 47 and K by T at position 49;
Q by N at position 48 and A by S at position 50;
Q by N at position 48 and A by T at position 50;
K by N at position 49 and E by S at position 51;
K by N at position 49 and E by T at position 51;
A by N at position 50 and T by S at position 52;
A by N at position 50 and T by T at position 52;
S by N at position 68 and K by S at position 70;
S by N at position 68 and K by T at position 70;
K by N at position 70 and S by S at position 72;
K by N at position 70 and S by T at position 72;
A by N at position 75 and D by S at position 77;
A by N at position 75 and D by T at position 77;
D by N at position 77 and T by S at position 79;
D by N at position 77 and T by T at position 79;
I by N at position 100 and G by S at position 102;

I by N at position 100 and G by T at position 102;
Q by N at position 101 and V by S at position 103;
Q by N at position 101 and V by T at position 103;
G by N at position 102 and G by S at position 104;
G by N at position 102 and G by T at position 104;
V by N at position 103 and V by S at position 105;
V by N at position 103 and V by T at position 105;
G by N at position 104 and T by S at position 106;
G by N at position 104 and T by T at position 106;
V by N at position 105 and E by S at position 107;
V by N at position 105 and E by T at position 107;
T by N at position 106 and T by S at position 108;
T by N at position 106 and T by T at position 108;
E by N at position 107 and P by S at position 109;
E by N at position 107 and P by T at position 109;
T by N at position 108 and I by S at position 110;
T by N at position 108 and I by T at position 110;
K by N at position 134 and S by S at position 136;
K by N at position 134 and S by T at position 136;
S by N at position 154 and N by S at position 156;
S by N at position 154 and N by T at position 156;
T by N at position 155 and L by S at position 157;
T by N at position 155 and L by T at position 157;
N by N at position 156 and Q by S at position 158;
N by N at position 156 and Q by T at position 158;
L by N at position 157 and E by S at position 159;
L by N at position 157 and E by T at position 159;
Q by N at position 158 and S by S at position 160;
Q by N at position 158 and S by T at position 160;
E by N at position 159 and L by S at position 161;
E by N at position 159 and L by T at position 161;

S by N at position 160 and R by S at position 162;
S by N at position 160 and R by T at position 162;
L by N at position 161 and S by S at position 163;
L by N at position 161 and S by T at position 163;
R by N at position 162 and K by S at position 164;
R by N at position 162 and K by T at position 164;
S by N at position 163 and E by S at position 165; and
S by N at position 163 and E by T at position 165,

where residue 1 corresponds to residue 1 of the mature IFN$\alpha$-2b or IFN$\alpha$-2a protein set forth in SEQ ID NO:1 or SEQ ID NO:182, respectively. In particular embodiments, the IFN$\alpha$-2b or IFN$\alpha$-2a mutant protein has increased resistance to proteolysis compared to unmodified IFN$\alpha$-2b or IFN$\alpha$-2a, and is selected from among proteins comprising one or more sets of duo-amino acid replacements in SEQ ID NO:1, corresponding to:

Q by N at position 5 and H by S at position 7;
P by N at position 39 and E by S at position 41;
P by N at position 39 and E by T at position 41;
Q by N at position 40 and E by S at position 42;
Q by N at position 40 and E by T at position 42;
E by N at position 41 and F by S at position 43;
E by N at position 41 and F by T at position 43;
F by N at position 43 and N by S at position 45;
G by N at position 44 and Q by T at position 46;
N by N at position 45 and F by S at position 47;
N by N at position 45 and F by T at position 47;
Q by N at position 46 and Q by S at position 48;
F by N at position 47 and K by S at position 49;
F by N at position 47 and K by T at position 49;
I by N at position 100 and G by S at position 102;
I by N at position 100 and G by T at position 102;
V by N at position 105 and E by S at position 107;

V by N at position 105 and E by T at position 107;
T by N at position 106 and T by S at position 108;
T by N at position 106 and T by T at position 108;
E by N at position 107 and P by S at position 109;
E by N at position 107 and P by T at position 109;
L by N at position 157 and E by S at position 159;
L by N at position 157 and E by T at position 159;
E by N at position 159 and L by S at position 161; and
E by N at position 159 and L by T at position 161.

**F. Protein Redesign**

Provided herein are methods for designing and generating new versions of native or modified cytokines, such as IFN$\alpha$-2b and IFN$\alpha$-2a. Using these methods, the redesigned cytokine maintains either sufficient, typically equal or improved levels of a selected phenotype, such as a biological activity, of the original protein, while at the same time its amino acid sequence is changed by replacement of up to: at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, at least 12%, at least 14%, at least 16%, at least 18%, at least 20%, at least 30%, at least 40% up to 50% or more of its native amino acids by the appropriate pseudo-wild type amino acids. Pseudo-wild type amino acids are those amino acids such that when they replace an original, such as native, amino acid at a given position on the protein sequence, the resulting protein displays substantially the same levels of biological activity (or sufficient activity for its therapeutic or other use) compared to the original, such as native, protein. In other embodiments, pseudo-wild type amino acids are those amino acids such that when they replace an original, such as native, amino acid at a given position on the protein sequence, the resulting protein displays the same phenotype, such as levels of biological activity, compared to an original, typically a native, protein. Pseudo-wild type amino acids and the appropriate replacing positions can be detected and

identified by any analytical or predictive means; such as for example, by performing an Alanine-scanning. Any other amino acid, particularly another amino acid that has a neutral effect on structure, such as Gly or Ser, also can be used for the scan. All those replacements of original, such as native, amino acids by Ala that do not lead to the generation of a HIT (a protein that has lost the desired biological activity), have either led to the generation of a LEAD (a protein with increased biological activity); or the replacement by Ala will be a neutral replacement, i.e., the resulting protein will display comparable levels of biological activity compared to the original, such as native, protein. The methods provided herein for protein redesign of cytokines, such as IFN$\alpha$-2b and IFN$\alpha$-2a, are intended to design and generate "artificial" (versus naturally existing) proteins, such that they consist of amino acid sequences not existing in Nature, but that display biological activities characteristic of the original, such as native, protein. These redesigned proteins are contemplated herein to be useful for avoiding potential side effects that might otherwise exist in other forms of cytokines in treatment of disease. Other uses of redesigned proteins provided herein are to establish cross-talk between pathways triggered by different proteins; to facilitate structural biology by generating mutants that can be crystallized while maintaining activity; and to destroy an activity of a protein without changing a second activity or multiple additional activities.

In one embodiment, a method for obtaining redesigned proteins includes *i*) identifying some or all possible target sites on the protein sequence that are susceptible to amino acid replacement without losing protein activity (protein activity in a largest sense of the term: enzymatic, binding, hormone, etc.) (These sites are the pseudo-wild type, Ψ-wt sites); *ii*) identifying appropriate replacing amino acids (Ψ-wt amino acids), specific for each Ψ-wt site, such that if used to replace the native amino acids at that specific Ψ-wt site, they can be expected to generate a protein with comparable biological activity compared to the original, such

as native, protein, thus keeping the biological activity of the protein substantially unchanged; *iii*) systematically introducing the specific Ψ-wt amino acids at every specific Ψ-wt position so as to generate a collection containing the corresponding mutant molecules. Mutants are generated, produced and phenotypically characterized one-by-one, in addressable arrays, such that each mutant molecule contains initially amino acid replacements at only one Ψ-wt site. In subsequent rounds mutant molecules also can be generated such that they contain one or more Ψ-wt amino acids at one or more Ψ-wt sites. Those mutant proteins carrying several mutations at a number of Ψ-wt sites, and that display comparable or improved biological activity are called redesigned proteins or Ψ-wt proteins. In particular embodiments, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, at least 15%, at least 20%, at least 25%, or more of the amino acid residue positions on a particular cytokine, such as IFN$\alpha$-2b and IFN$\alpha$-2a are replaced with an appropriate pseudo-wild type amino acid.

The first step is an amino acid scan over the full length of the protein. At this step, each and every one of the amino acids in the protein sequence is replaced by a selected reference amino acid, such as alanine. This permits the identification of "redesign-HIT" positions, i.e., positions that are sensitive to amino acid replacement. All of the other positions that are not redesign-HIT positions (i.e., those at which the replacement of the original, such as native, amino acid by the replacing amino acid, for example Ala, does not lead to a drop in protein fitness or biological activity) are referred to herein as "pseudo-wild type" positions. When the replacing amino acid, for example Ala, replaces the original, such as native, amino acid at a non-HIT position, then the replacement is neutral, in terms of protein activity, and the replacing amino acid is said to be a pseudo-wild type amino acid at that position. Pseudo-wild type positions appear to be less sensitive than redesign-HIT positions since

they tolerate the amino acid replacement without affecting the protein activity that is being either maintained or improved. Amino acid replacement at the pseudo-wild type positions, result in a non-change in the protein fitness (e.g., possess substantially the same biological activity), while at the same time to a divergence in the resulting protein sequence compared to the original, such as native, sequence.

To first identify those amino acid positions on the IFN$\alpha$-2b and IFN$\alpha$-2a protein that are involved or not involved in IFN$\alpha$-2b and IFN$\alpha$-2a protein activity, such as binding activity of IFN$\alpha$-2b and IFN$\alpha$-2a to its receptor, an Ala-scan was performed on the IFN$\alpha$-2b sequence as set forth in Example 4. For this purpose, each amino acid in the IFN$\alpha$-2b protein sequence was individually changed to Alanine. Any other amino acid, particularly another amino acid that has a neutral effect on structure, such as Gly or Ser, also can be used. Each resulting mutant IFN$\alpha$-2b protein was then expressed and the activity of the interferon molecule was then assayed. These particular amino acid positions, referred to herein as HITs would in principle not be suitable targets for amino acid replacement to increase protein stability, because of their involvement in the recognition of IFN-receptor or in the downstream pathways involved in IFN activity. For the Ala-scanning, the biological activity measured for the IFN$\alpha$-2b molecules was: *i*) their capacity to inhibit virus replication when added to permissive cells previously infected with the appropriate virus and, *ii*) their capacity to stimulate cell proliferation when added to the appropriate cells. The relative activity of each individual mutant compared to the native protein is indicated in FIG**10**A through C. HITs are those mutants that produce a decrease in the activity of the protein (in the example: all the mutants with activities below about 30% of the native activity.

In addition, the Alanine-scan was used to identify the amino acid residues on IFN$\alpha$-2b that when replaced with alanine correspond to 'pseudo-wild type' activity, i.e., those that can be replaced by alanine

without leading to a decrease in biological activity. Knowledge of these amino acids is useful for the re-design of the IFN$\alpha$-2b and IFN$\alpha$-2a proteins. The results are set forth in Table 5, and include pseudo-wild type amino acid positions of IFN$\alpha$-2b corresponding to SEQ ID NO:1, amino acid residues: 9, 10, 17, 20, 24, 25, 35, 37, 41, 52, 54, 56, 57, 58, 60, 63, 64, 65, 76, 89, and 90.

Accordingly, provided herein are IFN$\alpha$-2b and IFN$\alpha$-2a mutant proteins comprising one or more pseudo-wild type mutations at amino acid positions of IFN$\alpha$-2b or IFN$\alpha$-2a corresponding to SEQ ID NO:1 or SEQ ID NO:182, respectively, amino acid residues: 9, 10, 17, 20, 24, 25, 35, 37, 41, 52, 54, 56, 57, 58, 60, 63, 64, 65, 76, 89, and 90. The mutations can be either one or more of insertions, deletions and/or replacements of the native amino acid residue(s). In one embodiment, the pseudo-wild type replacements are mutations with alanine at each position. In another embodiment, the pseudo-wild type replacements are one or more mutations in SEQ ID NO:1 corresponding to:

L by A at position 9, L by A at position 17,
Q by A at position 20, I by A at position 24,
S by A at position 25, D by A at position 35,
G by A at position 37, E by A at position 41,
T by A at position 52, P by A at position 54,
L by A at position 56, H by A at position 57,
E by A at position 58, I by A at position 60,
I by A at position 63, F by A at position 64,
N by A at position 65, W by A at position 76,
Y by A at position 89, and Q by A at position 90.

In addition, the IFN$\alpha$-2b alanine scan revealed the following redesign-HITs having decreased antiviral activity at amino acid positions of IFN$\alpha$-2b corresponding to SEQ ID NO:1, amino acid residues: 2, 7, 8, 11, 13, 15, 16, 23, 26, 28, 29, 30, 31, 32, 33, 53, 69, 91, 93, 98, and 101. Accordingly, in particular embodiments where it is desired to

decrease the viral activity of IFN$\alpha$-2b or IFN$\alpha$-2a, either one or more of insertions, deletions and/or replacements of the native amino acid residue(s) can be carried out at one or more of amino acid positions of IFN$\alpha$-2b or IFN$\alpha$-2a corresponding to SEQ ID NO:1, amino acid residues: 2, 7, 8, 11, 13, 15, 16, 23, 26, 28, 29, 30, 31, 32, 33, 53, 69, 91, 93, 98, and 101.

Each of the redesign mutations set forth above can be combined with one or more of the IFN$\alpha$-2b or IFN$\alpha$-2a candidate LEAD mutations or one or more of the IFN$\alpha$-2b or IFN$\alpha$-2a LEAD mutants provided herein.

**G. 3D-scanning and Its Use for Modifying Cytokines**

Also provided herein is a method of structural homology analysis for comparing proteins regardless of their underlying amino acid sequences. For a subset of proteins families, such as the family of human cytokines, this information is rationally exploited to produce modified proteins. This method of structural homology analysis can be applied to proteins that are evolved by any method, including the 2D scanning method described herein. When used with the 2D method in which a particular phenotype, activity or characteristic of a protein is modified by 2D analysis, the method is referred to as 3D-scanning.

The use of "structural homology" analysis in combination with the directed evolution methods provided herein provides a powerful technique for identifying and producing various new protein mutants, such as cytokines, having desired biological activities, such as increased resistance to proteolysis. For example, the analysis of the "structural homology" between an optimized mutant version of a given protein and "structurally homologous" proteins allows identification of the corresponding structurally related or structurally similar amino acid positions (also referred to herein as "structurally homologous loci") on other proteins. This permits identification of mutant versions of the latter that have a desired optimized feature(s) (biological activity, phenotype) in a simple, rapid and predictive manner (regardless of amino acid sequence

and sequence homology). Once a mutant version of a protein is developed, then, by applying the rules of structural homology, the corresponding structurally related amino acid positions (and replacing amino acids) on other "structurally homologous" proteins readily are identified, thus allowing a rapid and predictive discovery of the appropriate mutant versions for the new proteins.

3-dimensionally structurally equivalent or similar amino acid positions that are located on two or more different protein sequences that share a certain degree of structural homology, have comparable functional tasks (activities and phenotypes). These two amino acids that occupy substantially equivalent 3-dimensional structural space within their respective proteins than can be said to be "structurally similar" or "structurally related" with each other, even if their precise positions on the amino acid sequences, when these sequences are aligned, do not match with each other. The two amino acids also are said to occupy "structurally homologous loci." "Structural homology" does not take into account the underlying amino acid sequence and solely compares 3-dimensional structures of proteins. Thus, two proteins can be said to have some degree of structural homology whenever they share conformational regions or domains showing comparable structures or shapes with 3-dimensional overlapping in space. Two proteins can be said to have a higher degree of structural homology whenever they share a higher amount of conformational regions or domains showing comparable structures or shapes with 3-dimensional overlapping in space. Amino acids positions on one or more proteins that are "structurally homologous" can be relatively far way from each other in the protein sequences, when these sequences are aligned following the rules of primary sequence homology. Thus, when two or more protein backbones are determined to be structurally homologous, the amino acid residues that are coincident upon three-dimensional structural superposition are referred to as "structurally similar" or "structurally related" amino acid

residues in structurally homologous proteins (also referred to as "structurally homologous loci"). Structurally similar amino acid residues are located in substantially equivalent spatial positions in structurally homologous proteins.

For example, for proteins of average size (approximately 180 residues), two structures with a similar fold will usually display rms deviations not exceeding 3 to 4 angstroms. For example, structurally similar or structurally related amino acid residues can have backbone positions less than 3.5, 3.0, 2.5, 2.0, 1.7 or 1.5 angstrom from each other upon protein superposition. RMS deviation calculations and protein superposition can be carried out using any of a number of methods known in the art. For example, protein superposition and RMS deviation calculations can be carried out using all peptide backbone atoms (e.g., N, C, C(C=O), O and CA (when present)). As another example, protein superposition can be carried out using just one or any combination of peptide backbone atoms, such as, for example, N, C, C(C=O), O and CA (when present). In addition, one skilled in the art will recognize that protein superposition and RMS deviation calculations generally can be performed on only a subset of the entire protein structure. For example, if the protein superposition is carried out using one protein that has many more amino acid residues than another protein, protein superposition can be carried out on the subset (e.g., a domain) of the larger protein that adopts a structure similar to the smaller protein. Similarly, only portions of other proteins can be suitable for superimposition. For example, if the position of the C-terminal residues from two structurally homologous proteins differ significantly, the C-terminal residues can be omitted from the structural superposition or RMS deviation calculations.

Accordingly, provided herein are methods of rational evolution of proteins based on the identification of potential target sites for mutagenesis (is-HITs) through comparison of patterns of protein backbone folding between structurally related proteins, irrespective of the

underlying sequences of the compared proteins. Once the structurally related amino acid positions are identified on the new protein, then suitable amino acid replacement criteria, such as PAM analysis, can be employed to identify candidate LEADs for construction and screening as described herein.

For example, analysis of "structural homology" between and among a number of related cytokines was used to identify on various members of the cytokine family, other than interferon alpha, those amino acid positions and residues that are structurally similar or structurally related to those found in the IFN$\alpha$-2b mutants provided herein that have been optimized for improved stability. The resulting modified cytokines are provided. This method can be applied to any desired phenotype using any protein, such as a cytokine, as the starting material to which an evolution procedure, such as the rational directed evolution procedure of U.S. application Serial No. 10/022,249 or the 2-dimensional scanning method provided herein, is applied. The structurally corresponding residues are then altered on members of the family to produce additional cytokines with similar phenotypic alterations.

**1) Homology**

Typically, homology between proteins is compared at the level of their amino acid sequences, based on the percent or level of coincidence of individual amino acids, amino acid per amino acid, when sequences are aligned starting from a reference, generally the residue encoded by the start codon. For example, two proteins are said to be "homologous" or to bear some degree of homology whenever their respective amino acid sequences show a certain degree of matching upon alignment comparison. Comparative molecular biology is primarily based on this approach. From the degree of homology or coincidence between amino acid sequences, conclusions can be made on the evolutionary distance between or among two or more protein sequences and biological systems.

The concept of "convergent evolution" is applied to describe the phenomena by which phylogenetically unrelated organisms or biological systems have evolved to share features related to their anatomy, physiology and structure as a response to common forces, constraints, and evolutionary demands from the surrounding environment and living organisms. Alternatively, "divergent evolution," is applied to describe the phenomena by which strongly phylogenetically related organisms or biological systems have evolved to diverge from identity or similarity as a response to divergent forces, constraints, and evolutionary demands from the surrounding environment and living organisms.

In the typical traditional analysis of homologous proteins there are two conceptual biases corresponding to: i) "convergent evolution," and ii) "divergent evolution." Whenever the aligned amino acid sequences of two proteins do not match well with each other, these proteins are considered "not related" or "less related" with each other and have different phylogenetic origins. There is no (or low) homology between these proteins and their respective genes are not homologous (or show little homology). If these two "non-homologous" proteins under study share some common functional features (e.g., interaction with other specific molecules, activity), they are determined to have arisen by "convergent evolution," i.e., by evolution of their non-homologous amino acid sequences, in such a way that they end up generating functionally "related" structures.

On the other hand, whenever the aligned amino acid sequences of two proteins do match with each other to a certain degree, these proteins are considered to be "related" and to share a common phylogenetic origin. A given degree of homology is assigned between these two proteins and their respective genes likewise share a corresponding degree of homology. During the evolution of their initial highly homologous amino acid sequence, enough changes can be accumulated in such a way that they end up generating "less-related" sequences and less related

function. The divergence from perfect matching between these two "homologous" proteins under study is said come from "divergent evolution."

**2) 3D-scanning (Structural Homology) methods**

Structural homology refers to homology between the topology and three-dimensional structure of two proteins. Structural homology is not necessarily related to "convergent evolution" or to "divergent evolution," nor is it related to the underlying amino acid sequence. Rather, structural homology is likely driven (through natural evolution) by the need of a protein to fit specific conformational demands imposed by its environment. Particular structurally homologous "spots" or "loci" would not be allowed to structurally diverge from the original structure, even when its own underlying sequence does diverge. This structural homology is exploited herein to identify loci for mutation.

Within the amino acid sequence of a protein resides the appropriate biochemical and structural signals to achieve a specific spatial folding in either an independent or a chaperon-assisted manner. Indeed, this specific spatial folding ultimately determines protein traits and activity. Proteins interact with other proteins and molecules in general through their specific topologies and spatial conformations. In principle, these interactions are not based solely on the precise amino acid sequence underlying the involved topology or conformation. If protein traits, activity (behavior and phenotypes) and interactions rely on protein topology and conformation, then evolutionary forces and constraints acting on proteins can be expected to act on topology and conformation. Proteins sharing similar functions will share comparable characteristics in their topology and conformation, despite the underlying amino acid sequences that create those topologies and conformations.

**3) Application of 3D Scanning to Cytokines**

The method based on structural homology, including the 3D-scanning method provided herein can be applied to any related proteins.

For exemplary purposes herein it is applied to cytokines. In exemplary embodiments, methods for altering phenotypes of members of families of cytokines by altering one member such as by employing the 2-dimensional rational scanning method are provided. As provided herein, other members of these cytokine families then can be similarly modified by identifying and changing structurally homologous residues to similarly alter the phenotypes of such proteins.

In an exemplary embodiment herein, IFN$\alpha$-2b mutants with increased resistance to proteolysis are generated by the 2-dimensional rational scanning method; IFN$\beta$ mutants also were generated. The corresponding residues on members of cytokine families that possess structural homology to IFN$\alpha$-2b were identified and the identifeid residues on the other cytokines were similarly modified to produce cytokines with increased resistance to proteolysis. Hence also provided herein are cytokine mutants that display increased resistance to proteolysis and/or glomerular filtration containing one or more amino acid replacements.

Provided herein are mutant (modified) cytokines that display altered features and properties, such as a resistance to proteolysis. Methods for producing such modified cytokines also are provided.

Also provided herein is a method of structural homology analysis for comparing proteins regardless their underlying amino acid sequences. For a subset of proteins families, such as the family of human cytokines, this information is rationally exploited herein. Human cytokines all share a common helix bundle fold, which is used to structurally define the 4-helical cytokine superfamily in the structural classification of the protein database SCOP© (Structural Classification of Proteins; see, e.g., Murzin *et al.*, *J. Mol. Biol.*, 247:536-540, 1995 and "http://scop.mrc-lmb.cam.ac.uk/scop/"). This superfamily includes three different families: 1) the interferons/interleukin-10 protein family (SEQ ID NOS: 1 and 182-200); 2) the long-chain cytokine family (SEQ ID NOS: 210-217); and 3) the short-chain cytokine family (SEQ ID NOS: 201-209).

For example, a distinct feature of cytokines from the interferons/interleukin-10 family is an additional (fifth) helix. This family includes interleukin-10 (IL-10; SEQ ID NO:200, interferon beta (IFN$\beta$; SEQ ID NO: 196), interferon alpha-2a (IFN$\alpha$-2a; SEQ ID NO: 182), interferon alpha-2b (IFN$\alpha$-2b; SEQ ID NO:1), and interferon gamma (IFN-$\gamma$; SEQ ID NO: 199). The long-chain cytokine protein family includes, among others, granulocyte colony stimulating factor (G-CSF; SEQ ID NO: 210), leukemia inhibitory factor (LIF; SEQ ID NO: 213), growth hormone (hGH; SEQ ID NO: 216), ciliary neurotrophic factor (CNTF; SEQ ID NO: 212), leptin (SEQ ID NO: 211), oncostatin M (SEQ ID NO: 214), interleukin-6 (IL-6; SEQ ID NO: 217) and interleukin-12 (IL-12; SEQ ID NO: 215). The short-chain cytokine protein family includes, among others, erythropoietin (EPO; SEQ ID NO: 201), granulocyte-macrophage colony stimulating factor (GM-CSF; SEQ ID NO: 202), interleukin-2 (IL-2; SEQ ID NO: 204), interleukin-3 (IL-3; SEQ ID NO: 205), interleukin-4 (IL-4; SEQ ID NO: 207), interleukin-5 (IL-5; SEQ ID NO: 208), interleukin-13 (IL-13; SEQ ID NO: 209), Flt3 ligand (SEQ ID NO: 203) and stem cell factor (SCF; SEQ ID NO: 206).

Although the degree of similarity among the underlying amino acid sequences of these cytokines does not appear high, their corresponding 3-dimensional structures present a high level of similarity (see, e.g., FIGS8B through D). Effectively, the best structural similarity is obtained between two 3-dimensional protein structures of the same family in the 4-helical cytokine superfamily.

The methods provided herein for producing mutant cytokines are exemplified with reference to production of cytokines that display a substantially equivalent increase in resistance to proteolysis relative to the optimized IFN$\alpha$-2b mutants. It is understood that this method can be applied to other families of proteins and for other phenotypes.

In one embodiment, proteins of the 4-helical cytokine superfamily are provided herein that are structurally homologous IFN$\alpha$-2b LEAD mutants set forth herein. For example, by virtue of the knowledge of the

3-dimensional structural amino acid positions within the LEAD IFN$\alpha$-2b mutants provided herein that confer higher resistance to a challenge with either proteases or blood lysate or serum, while maintaining or improving the requisite biological activity, the corresponding structurally related (e.g., structurally similar) amino acid residues on a variety of other cytokines are identified (FIG**9**).

Numerous methods are well known in the art for identifying structurally related amino acid positions with 3-dimensionally structurally homologous proteins. Exemplary methods include, but are not limited to: CATH (Class, Architecture, Topology and Homologous superfamily) which is a hierarchical classification of protein domain structures based on four different levels (Orengo *et al.*, *Structure*, 5(8):1093-1108, 1997); CE (Combinatorial Extension of the optimal path), which is a method that calculates pairwise structure alignments (Shindyalov *et al.*, *Protein Engineering*, 11(9):739-747, 1998); FSSP (Fold classification based on Structure-Structure alignment of Proteins), which is a database based on the complete comparison of all 3-dimensional protein structures that currently reside in the Protein Data Bank (PDB) (Holm *et al.*, *Science*, 273:595-602, 1996); SCOP (Structural Classification of Proteins), which provides a descriptive database based on the structural and evolutionary relationships between all proteins whose structure is known (Murzin *et al.*, *J. Mol. Biol.*, 247:536-540, 1995); and VAST (Vector Alignment Search Tool), which compares newly determined 3-dimensional protein structure coordinates to those found in the MMDB/PDB database (Gibrat *et al.*, *Current Opinion in Structural Biology*, 6:377-385, 1995).

In an exemplary embodiment, the step-by-step process including the use of a program referred to as TOP (see FIG**8**A and Lu, G., *J. Appl. Cryst.*, 33:176-189, 2000; publicly available, for example, at bioinfo1.mbfys.lu.se/TOP is used for protein structure comparison. This program runs two steps for each protein structure comparison. In the first step topology of secondary structure in the two structures is

compared. The program uses two points to represent each secondary structure element (alpha helices or beta strands) then systematically searches all the possible super-positions of these elements in 3-dimensional space (defined as the root mean square deviation – rmsd, the angle between the two lines formed by the two points and the line-line distance). The program searches to determine whether additional secondary structure elements can fit by the same superposition operation. If secondary structures that can fit each other exceed a given number, the program identifies the two structures as similar. The program gives as an output a comparison score called "Structural Diversity" that considers the distance between matched $\alpha$-carbon atoms and the number of matched residues. The lower the "Structural Diversity" score, the more the two structures are similar. In various embodiments herein, the Structural Diversity scores range from 0 up to about 67.

In the exemplified embodiment, all the cytokines were first structurally aligned against the IFN$\alpha$-2b structure. For the proteins within the same family as IFN$\alpha$-2b (e.g., the interferons/interleukin-10 cytokine family), this alignment was directly used to identify the structurally related is-HIT target amino acid positions and/or regions corresponding to the structurally homologous positions and/or regions on IFN$\alpha$-2b where LEAD mutants were found (FIG**8**B). For the other cytokines, the protein of the family (either long- or short-chain cytokines) with the best 3-dimensional structural alignment with IFN$\alpha$-2b was selected using the lowest "Structural Diversity" score as the representative for that family. From the short-chain cytokine protein family, erythropoietin (EPO; see FIG**8**C) was identified as the best structural homologue of IFN$\alpha$-2b (rmsd = 1.9 angstroms; number of aligned residues = 62; Structural Diversity = 13.8). From the long-chain cytokine protein family, granulocyte-colony stimulating factor (G-CSF; see FIG**8**D) was identified as the best structural homologue of IFN$\alpha$-2b (rmsd = 1.7 angstroms; number of aligned residues = 77; Structural Diversity = 7.8).

Next, the amino acid positions and/or regions corresponding to the LEAD mutant regions on IFN$\alpha$-2b were identified on these two proteins. These two best structural homologues of IFN$\alpha$-2b (e.g., EPO and G-CSF; see FIGS**12**L and **12**E, respectively) were structurally aligned to each of the other cytokines within their respective cytokine protein families. As a result, protein regions likely to be targets for serum protease resistance were identified on all cytokines (see FIGS**12**A through T). Amino acids in these target regions were then checked for their exposure to the solvent and their susceptibility to be protease substrate. Exposed and substrate residues are then subjected to PAM250 analysis as set forth above, so that a group of non-substrate and functionally conservative amino acid residues are selected as replacements. The results of the above structural homology analysis for each of the cytokines provided herein are set forth in FIGS**12**A through T.

Accordingly, provided herein are modified cytokines that exhibit greater resistance to proteolysis compared to the unmodified cytokine protein, comprising one or more amino acid replacements at one or more target positions on the cytokine corresponding to a structurally-related modified amino acid position within the 3-dimensional structure of an IFN$\alpha$-2b modified protein provided herein. The resistance to proteolysis can be measured by mixing it with a protease in vitro, incubation with blood or incubation with serum. Also provided herein are cytokine structural homologues of an IFN$\alpha$-2b modified protein provided herein, comprising one or more amino acid replacements in the cytokine structural homologue at positions corresponding to the 3-dimensional-structurally-similar modified positions within the 3-dimensional structure of the modified IFN$\alpha$-2b. In one embodiment, the cytokine homologue has increased resistance to proteolysis compared to its unmodified and/or wild type cytokine counterpart. Resistance to proteolysis can be measured by mixture with a protease in vitro, incubation with blood, or incubation with serum.

**a) Structurally Homologous Interferon Mutants**

Also provided herein are modified cytokines or cytokine structural homologues of IFNα-2b that are IFNα cytokines. These IFNα cytokines include, but are not limited to, IFNα-2a, IFNα-c, IFNα-2c, IFNα-d, IFNα-5, IFNα-6, IFNα-4, IFNα-4b, IFNα-I, IFNα-J, IFNα-H, IFNα-F, IFNα-8 and IFNα-consensus cytokine (see, SEQ ID No. 232). Accordingly, amont the the modified IFNα cytokines provided herein are those with one or more amino acid replacements at one or more target positions in either IFNα-2a, IFNα-c, IFNα-2c, IFNα-d, IFNα-5, IFNα-6, IFNα-4, IFNα-4b, IFNα-I, IFNα-J, IFNα-H, IFNα-F, IFNα-8, or IFNα-consensus cytokine corresponding to a structurally-related modified amino acid position within the 3-dimensional structure of the IFNα-2b modified proteins provided herein The replacements lead to greater resistance to proteases, as assessed by incubation with a protease or a with a blood lysate or by incubation with serum, compared to the unmodified IFN alpha-2a.

In particular embodiments, the modified IFNα cytokines are selected from among:

the modified IFNα-2a that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 182, corresponding to amino acid positions: 41, 58, 78, 107, 117, 125, 133 and 159;

the modified IFNα-c that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 183, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the modified IFNα-2c cytokine that is human and is selected from among cytokines comprising one or more single amino acid replacements in SEQ ID NO: 185, corresponding to amino acid positions: 41, 58, 78, 107, 117, 125, 133 and 159;

the IFNα-d modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in

SEQ ID NO: 186, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-5 modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 187, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-6 modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 188, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-4 modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 189, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-4b modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 190, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-I modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 191, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-J modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 192, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-H modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 193, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-F modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 194, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160;

the IFN$\alpha$-8 modified protein that is human and is selected from among proteins comprising one or more single amino acid replacements in SEQ ID NO: 195, corresponding to amino acid positions: 41, 59, 79, 108, 118, 126, 134 and 160; and

the IFN$\alpha$-consensus modified protein that is human and is selected from among proteins that contain one or more single amino acid replacements in SEQ ID NO: 232, corresponding to amino acid positions: 41, 58, 78, 107, 117, 125, 133 and 159.

**b) Structurally Homologous Cytokine Mutants**

As set forth above, provided herein are modified cytokines that contain one or more amino acid replacements at one or more target positions in either interleukin-10 (IL-10), interferon beta (IFN$\beta$), IFN$\beta$-1, IFN$\beta$-2a, interferon gamma (IFN-$\gamma$), granulocyte colony stimulating factor (G-CSF), and human erythropoietin (EPO); corresponding to a structurally-related modified amino acid position within the 3-dimensional structure of the IFN$\alpha$-2b modified proteins provided herein. The replacements lead to greater resistance to proteases, as assessed by incubation with a protease or a with a blood lysate or by incubation with serum, compared to the unmodified cytokine.

Also provided herein are modified cytokines that contain one or more amino acid replacements at one or more target positions in either granulocyte-macrophage colony stimulating factor (GM-CSF), interleukin-2 (IL-2), interleukin-3 (IL-3), interleukin-4 (IL-4), interleukin-5 (IL-5), interleukin-13 (IL-13), Flt3 ligand and stem cell factor (SCF); corresponding to a structurally-related modified amino acid position within the 3-dimensional structure of the human EPO modified proteins provided herein. The replacements lead to greater resistance to proteases, as

assessed by incubation with a protease or a with a blood lysate or by incubation with serum, compared to the unmodified cytokine.

Also provided herein are modified cytokines that contain one or more amino acid replacements at one or more target positions in either interleukin-10 (IL-10), interferon beta (IFN$\beta$), interferon gamma (IFN-$\gamma$), human granulocyte colony stimulating factor (G-CSF), leukemia inhibitory factor (LIF), human growth hormone (hGH), ciliary neurotrophic factor (CNTF), leptin, oncostatin M, interleukin-6 (IL-6) and interleukin-12 (IL-12); corresponding to a structurally-related modified amino acid position within the 3-dimensional structure of the human G-CSF modified proteins provided herein. The replacements lead to greater resistance to proteases, as assessed by incubation with a protease or a with a blood lysate or by incubation with serum, compared to the unmodified cytokine.

In particular embodiments, the modified cytokines are selected from the following.

A modified IFN$\beta$ cytokine, comprising mutations at one or more amino acid residues of IFN$\beta$ corresponding to SEQ ID NO: 196: 39, 42, 45, 47, 52, 67, 71, 73, 81, 107, 108, 109, 110, 111, 113, 116, 120, 123, 124, 128, 130, 134, 136, 137, 163 and 165. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In particular embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 196 set forth in FIG**12**A corresponding to SEQ ID NOS: 233-289, where the first amino acid indicated is substituted by the second at the position indicated for all of the substitutions set forth in FIG**12**A through T.

A modified IFN$\beta$-1 cytokine, comprising mutations at one or more amino acid residues of IFN$\beta$-1 corresponding to SEQ ID NO: 197: 39, 42, 45, 47, 52, 67, 71, 73, 81, 107, 108, 109, 110, 111, 113, 116, 120, 123, 124, 128, 130, 134, 136, 137, 163 and 165. The mutations include insertions, deletions and replacements of the native amino acid residue(s).

A modified IFN$\beta$-2a cytokine, comprising mutations at one or more amino acid residues of IFN$\beta$-2a corresponding to SEQ ID NO:198: 39, 42, 45, 47, 52, 67, 71, 73, 81, 107, 108, 109, 110, 111, 113, 116, 120, 123, 124, 128, 130, 134, 136, 137, 163 and 165. The mutations include insertions, deletions and replacements of the native amino acid residue(s).

A modified IFN-gamma cytokine, comprising mutations at one or more amino acid residues of IFN-gamma corresponding to SEQ ID NO:199: 33, 37, 40, 41, 42, 58, 61, 64, 65 and 66. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In particular embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO:199 set forth in FIG**12**B corresponding to SEQ ID NOS: 290-311.

A modified IL-10 cytokine, comprising mutations at one or more amino acid residues of IL-10 corresponding to SEQ ID NO:200: 49, 50, 52, 53, 54, 55, 56, 57, 59, 60, 67, 68, 71, 72, 74, 75, 78, 81, 84, 85, 86, and 88. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, replacements are selected from among amino acid substitutions in SEQ ID NO:200 set forth in FIG**12**C corresponding to SEQ ID NOS: 312-361.

A modified erythropoietin cytokine, comprising mutations at one or more amino acid residues of erythropoietin corresponding to SEQ ID NO:201: 43, 45, 48, 49, 52, 53, 55, 72, 75, 76, 123, 129, 130, 131, 162, and 165. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 201 set forth in FIG**12**L corresponding to SEQ ID NOS: 940-977.

A modified GM-CSF cytokine, comprising mutations at one or more amino acid residues of GM-CSF corresponding to SEQ ID NO: 202: 38, 41, 45, 46, 48, 49, 51, 60, 63, 67, 92, 93, 119, 120, 123, and 124.

The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 202 set forth in FIG**12**N corresponding to SEQ ID NOS: 362-400.

A modified Flt3 ligand cytokine, comprising mutations at one or more amino acid residues of Flt3 ligand corresponding to SEQ ID NO: 203: 3, 40, 42, 43, 55, 58, 59, 61, 89, 90, 91, 95, and 96. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 203 set forth in FIG**12**M corresponding to SEQ ID NOS: 401-428.

A modified IL-2 cytokine, comprising mutations at one or more amino acid residues of IL-2 corresponding to SEQ ID NO: 204 at positions 43, 45, 48, 49, 52, 53, 60, 61, 65, 67, 68, 72, 100, 103, 104, 106, 107, 109, 110, and 132. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 204 set forth in FIG**12**P and SEQ ID NOS: 429-476.

A modified IL-3 cytokine, comprising mutations at one or more amino acid residues of IL-3 corresponding to SEQ ID NO: 205: 37, 43, 46, 59, 63, 66, 96, 100, 101, and 103. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO:205 set forth in FIG**12**Q corresponding to SEQ ID NOS: 477-498.

A modified SCF cytokine, comprising mutations at one or more amino acid residues of SCF corresponding to SEQ ID NO: 206: 27, 31, 34, 37, 54, 58, 61, 62, 63, 96, 98, 99, 100, 102, 103, 106, 107, 108, 109, 134, and 137. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary

embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 206 set forth in FIG12T corresponding to SEQ ID NOS: 499-542.

A modified IL-4 cytokine, comprising mutations at one or more amino acid residues of IL-4 corresponding to SEQ ID NO: 207: 26, 37, 53, 60, 61, 64, 66, 100, 102, 103, and 126. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 207 set forth in FIG12R corresponding to SEQ ID NOS: 543-567.

A modified IL-5 cytokine, comprising mutations at one or more amino acid residues of IL-5 corresponding to SEQ ID NO: 208: 32, 34, 39, 46, 47, 56, 84, 85, 88, 89, 90, 102, 110, and 111. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 208 set forth in FIG12S corresponding to SEQ ID NOS: 568-602.

A modified IL-13 cytokine, comprising mutations at one or more amino acid residues of IL-13 corresponding to SEQ ID NO: 209: 32, 34, 38, 48, 79, 82, 85, 86, 88, 107, 108, 110, and 111. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 209 set forth in FIG12O corresponding to SEQ ID NOS: 603-630.

A modified G-CSF cytokine, comprising mutations at one or more amino acid residues of G-CSF corresponding to SEQ ID NO: 210: 61, 63, 68, 72, 86, 96, 100, 101, 131, 133, 135, 147, 169, 172, and 177. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 210 set forth in FIG12E corresponding to SEQ ID NOS: 631-662.

A modified leptin cytokine, comprising mutations at one or more amino acid residues of leptin corresponding to SEQ ID NO: 211: 43, 49, 99, 100, 104, 105, 107, 108, 141 and 142. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 211 set forth in FIG**12**I corresponding to SEQ ID NOS: 663-683.

A modified CNTF cytokine, comprising mutations at one or more amino acid residues of CNTF corresponding to SEQ ID NO: 212: 62, 64, 66, 67, 86, 89, 92, 100, 102, 104, 131, 132, 133, 135, 136, 138, 140, 143, 148, and 151. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 212 set forth in FIG**12**D corresponding to SEQ ID NOS: 684-728.

A modified LIF cytokine, comprising mutations at one or more amino acid residues of LIF corresponding to SEQ ID NO: 213: 69, 70, 85, 99, 102, 104, 106, 109, 137, 143, 146, 148, 149, 153, 154, and 156. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 213 set forth in FIG**12**J corresponding to SEQ ID NOS: 729-760.

A modified oncostatin M cytokine, comprising mutations at one or more amino acid residues of oncostatin M corresponding to SEQ ID NO: 214: 59, 60, 63, 65, 84, 87, 89, 91, 94, 97, 99, 100, 103, and 106. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 214 set forth in FIG**12**K corresponding to SEQ ID NOS: 761-793.

A modified IL-12 cytokine, comprising mutations at one or more amino acid residues of IL-12 corresponding to SEQ ID NO: 215: 56, 61,

66, 67, 68, 70, 72, 75, 78, 79, 82, 89, 92, 93, 107, 110, 111, 115, 117, 124, 125, 127, 128, 129, and 189. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 215 set forth in FIG**12**G corresponding to SEQ ID NOS: 794-849.

A modified hGH cytokine, comprising mutations at one or more amino acid residues of hGH corresponding to SEQ ID NO: 216: 56, 59, 64, 65, 66, 88, 92, 94, 101, 129, 130, 133, 134, 140, 143, 145, 146, 147, 183, and 186. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 216 set forth in FIG**12**F corresponding to SEQ ID NOS: 850-895.

A modified IL-6 cytokine, comprising mutations at one or more amino acid residues of IL-6 corresponding to SEQ ID NO: 217: 64, 65, 66, 68, 69, 75, 77, 92, 98, 103, 105, 108, 133, 138, 139, 140, 149, 156, 178, and 181. The mutations include insertions, deletions and replacements of the native amino acid residue(s). In exemplary embodiments, the replacements are selected from among amino acid substitutions in SEQ ID NO: 217 set forth in FIG**12**H corresponding to SEQ ID NOS: 896-939.

In certain embodiments, the modified cytokines provided herein possess increased stability compared to the unmodified cytokine. Stability can be assessed by any *in vitro* or *in vivo* method, such as by measuring residual inhibition of viral replication or to stimulation of cell proliferation in appropriate cells, after incubation with either mixtures of proteases, individual proteases, blood lysate or serum.

In other embodiments, the modified cytokines provided herein possess decreased stability compared to the unmodified cytokine. Stability can be assessed by any *in vitro* or *in vivo* method, such as by

measuring residual inhibition of viral replication or to stimulation of cell proliferation in appropriate cells, after incubation with either mixtures of proteases, individual proteases, blood lysate or serum.

In ther embodiments, the modified cytokines provided herein possess increased activity compared to the unmodified cytokine. Stability can be assessed by any *in vitro* or *in vivo* method, such as by measuring residual inhibition of viral replication or to stimulation of cell proliferation in appropriate cells, after incubation with either mixtures of proteases, individual proteases, blood lysate or serum.

**H. Rational evolution of IFNβ for increased resistance to proteolysis and/or higher conformational stability**

Treatment with interferon b (IFNβ) is a well established therapy. Typically it is used for treatment of multiple sclerosis (MS). Patients receiving interferon β are subject to frequent repeat applications of the drug. The instability of IFNβ in the blood stream and under the storage conditions is well known. Hence it would be useful to increasing stability (half-life) of IFNβ in serum and also *in vitro* would improve it as a drug.

The 2D-scanning method and the 3D-scanning method (using structural homology) provided herein (see, copending ) were each applied to interferon β. Provided herein are mutant variants of the IFNβ protein that display improved stability as assessed by resistance to proteases (thereby possessing increased protein half-life) and at least comparable biological activity as assessed by antiviral or antiproliferation activity compared to the unmodified and wild type native IFNβ protein (SEQ ID NO: 196). The IFNβ mutant proteins provided herein confer a higher half-life and at least comparable biological activity with respect to the native sequence. Thus, the optimized IFNβ protein mutants provided herein that possess increased resistance to proteolysis result in a decrease in the frequency of injections needed to maintain a sufficient drug level in serum, thus leading to, for example: *i*) higher comfort and acceptance by patients, *ii*) lower doses necessary to achieve comparable biological

effects, and *iii*) as a consequence of (*ii*), likely attenuation of any secondary effects.

In exemplary embodiments, the half-life of the IFN$\beta$ mutants provided herein is increased by an amount selected from at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500% or more, when compared to the half-life of native human IFN$\beta$ in either human blood, human serum or an in vitro mixture containing one or more proteases. In other embodiments, the half-life of the IFN$\beta$ mutants provided herein is increased by an amount selected from at least 6 times, 7 times, 8 times, 9 times, 10 times, 20 times, 30 times, 40 times, 50 times, 60 times, 70 times, 80 times, 90 times, 100 times, 200 times, 300 times, 400 times, 500 times, 600 times, 700 times, 800 times, 900 times, 1000 times, or more, when compared to the half-life of native human IFN$\beta$ in either human blood, human serum or an in vitro mixture containing one or more proteases.

Two approaches were used herein to increase the stability of IFN$\beta$ by amino acid replacement: *i*) Resistance to proteases: amino acid replacement that leads to higher resistance to proteases by direct destruction of the protease target residue or sequence, while either maintaining or improving the requisite biological activity (e.g., antiviral and anti-proliferation activity), and/or *ii*) Conformational stability: amino acid replacement that leads to an increase in conformational stability (i.e. half-life at room temperature or at 37°C), while either improving or maintaining the requisite biological activity (e.g., antiviral and anti-proliferation activity).

Two methodologies were used to address the improvements described above: (a) 2D-scanning methods were used to identify aminoacid changes that lead to improvement in protease resistance and to improvement in conformational stability, and (b) 3D-scanning, which

employs structural homology methods methods also were used to identify aminoacid changes that lead to improvement in protease resistance. The 2D-scanning and 3D-scanning methods each were used to identify the amino acid changes on IFN$\beta$ that lead to an increase in stability when challenged either with proteases, human blood lysate or human serum. Increasing protein stability to proteases, human blood lysate or human serum is contemplated herein to provide a longer *in vivo* half-life for the particular protein molecules, and thus a reduction in the frequency of necessary injections into patients. The biological activities that have been measured for the IFN$\beta$ molecules are *i*) their capacity to inhibit virus replication when added to permissive cells previously infected with the appropriate virus, and *ii*) their capacity to stimulate cell proliferation when added to the appropriate cells. Prior to the measurement of biological activity, IFN$\beta$ molecules were challenged with proteases, human blood lysate or human serum during different incubation times. The biological activity measured, corresponds then to the residual biological activity following exposure to the proteolytic mixtures.

As set forth above, provided herein are methods for the generating IFN$\beta$ molecules (or any target protein, particularly cytokines) that, while maintaining a requisite biological activity without substantial change (sufficient for therapeutic application(s)), have been rendered less susceptible to digestion by blood proteases and therefore display a longer half-life in blood circulation. In this particular example, the method used included the following specific steps as exemplified in the Examples: For the improvement of resistance to proteases, by 2D-scanning, the method included:

1) Identifying some or all possible target sites on the protein sequence that are susceptible to digestion by one or more specific proteases (these sites are the is-HITs); and

2) Identifying appropriate replacing amino acids, specific for each is-HIT, such that if used to replace one or more of the original amino acids

at that specific is-HIT, they can be expected to increase the is-HIT's resistance to digestion by protease while at the same time, keeping the biological activity of the protein unchanged (these replacing amino acids are the candidate LEADs).

For the improvement of resistance to proteases, by 3D-scanning (structural homology):

1) Identifying some or all possible target sites (is-HITS) on the protein sequence that display an acceptable degree of structural homology around the aminoacid positions mutated in the LEAD molecules previously obtained for IFN$\alpha$ using 2D-scanning, and that are susceptible to digestion by one or more specific proteases; and

2) Identifying appropriate replacing amino acids, specific for each is-HIT, such that if used to replace one or more of the original amino acids at that specific is-HIT, they can be expected to increase the is-HIT's resistance to digestion by protease while at the same time, keeping the biological activity of the protein unchanged (these replacing amino acids are the candidate LEADs).

For the improvement of conformational stability, by 2D-scanning, as provided herein:

1) Identifying some or all possible target sites on the protein sequence that are susceptible to being directly involved in the intramolecular flexibility and conformational change (these sites are the is-HITs); and

2) Identifying appropriate replacing amino acids, specific for each is-HIT, such that if used to replace one or more of the original amino acids at that specific is-HIT, they can be expected to increase the thermal stability of the molecule while at the same time, keeping the biological activity of the protein unchanged (these replacing amino acids are the candidate LEADs).

See Figures 6(O)-6(S) and Figure 8(A).

Using the 2D-scanning and 3D-scanning methods and the 3-dimensional structure of IFN$\beta$, the following amino acid target positions were identified as is-HITs on IFN$\beta$, which numbering is that of the mature protein (SEQ ID NO:196):

**By 3D-scanning** (see, SEQ ID Nos. 234-289, 989-1015): D by Q at position 39, D by H at position 39, D by G at position 39, E by Q at position 42, E by H at position 42, K by Q at position 45, K by T at position 45, K by S at position 45, K by H at position 45, L by V at position 47, L by I at position 47, L by T at position 47, L by Q at position 47, L by H at position 47, L by A at position 47, K by Q at position 52, K by T at position 52, K by S at position 52, K by H at position 52, F by I at position 67, F by V at position 67, R by H at position 71, R by Q at position 71, D by H at position 73, D by G at position 73, D by Q at position 73, E by Q at position 81, E by H at position 81, E by Q at position 107, E by H at position 107, K by Q at position 108, K by T at position 108, K by S at position 108, K by H at position 108, E by Q at position 109, E by H at position 109, D by Q at position 110, D by H at position 110, D by G at position 110, F by I at position 111, F by V at position 111, R by H at position 113, R by Q at position 113, L by V at position 116, L by I at position 116, L by T at position 116, L by Q at position 116, L by H at position 116, L by A at position 116, L by V at position 120, L by I at position 120, L by T at position 120, L by Q at position 120, L by H at position 120, L by A at position 120, K by Q at position 123, K by T at position 123, K by S at position 123, K by H at position 123, R by H at position 124,, R by Q at position 124, R by H at position 128, R by Q at position 128, L by V at position 130, L by I at position 130, L by T at position 130, L by Q at position 130, L by H at position 130, L by A at position 130, K by Q at position 134, K by T at position 134, K by S at position 134, K by H at position 134, K by Q at position 136, K by T at position 136, K by S at position 136,, K by H at position 136, E by Q at position 137, E by H at

position 137, Y by H at position 163, Y by I at position 163I, R by H at position 165, R by Q at position 165.

**By 2D-scanning (see SEQ ID Nos.1016-1302)** : M by V at position 1, M by I at position 1, M by T at position 1, M by Q at position 1, M by A at position 1, L by V at position 5, L by I at position 5, L by T at position 5, L by Q at position 5, L by H at position 5, L by A at position 5, F by I at position 8, F by V at position 8, L by V at position 9, L by I at position 9, L by T at position 9, L by Q at position 9, L by H at position 9, L by A at position 9, R by H at position 11, R by Q at position 11, F by I at position 15, F by V at position 15, K by Q at position 19, K by T at position 19, K by S at position 19, K by H at position 19, W by S at position 22, W by H at position 22, N by H at position 25, N by S at position 25, N by Q at position 25, R by H position 27, R by Q position 27, L by V at position 28, L by I at position 28, L by T at position 28, L by Q at position 28, L by H at position 28, L by A at position 28, E by Q at position 29, E by H at position 29, Y by H at position 30, Y by I at position 30, L by V at position 32, L by I at position 32, L by T at position 32, L by Q at position 32, L by H at position 32, L by A at position 32, K by Q at position 33, K by T at position 33, K by S at position 33, K by H at position 33, R by H at position 35, R by Q at position 35, M by V at position 36, M by I at position 36, M by T at position 36, M by Q at position 36, M by A at position 36, D by Q at position 39, D by H at position 39, D by G at position 39, E by Q at position 42, E by H at position 42, K by Q at position 45, K by T at position 45, K by S at position 45, K by H at position 45, L by V at position 47, L by I at position 47, L by T at position 47, L by, Q at position 47, L by H at position 47, L by A at position 47, K by Q at position 52, K by T at position 52, K by S at position 52, K by H at position 52, F by I at position 67, F by V at position 67, R by H at position 71, R by Q at position 71, D by Q at position 73, D by H at position 73, D by G at position 73, E by Q at position 81, E by H at

position 81, E by Q at position 85, E by H at position 85, Y by H at position 92, Y by I at position 92, K by Q at position 99, K by T at position 99, K by S at position 99, K by H at position 99, E by Q at position 103, E by H at position 103, E by Q at position 104, E by H at position 104, K by Q at position 105, K by T at position 105, K by S at position 105, K by H at position 105, E by Q at position 107, E by H at position 107, K by Q at position 108, K by T at position 108, K by S at position 108, K by H at position 108, E by Q at position 109, E by H at position 109, D by Q at position 110, D by H at position 110, D by G at position 110, F by I at position 111, F by V at position 111, R by H at position 113, R by Q at position 113, L by V at position 116, L by I at position 116, L by T at position 116, L by Q at position 116, L by H at position 116, L by A at position 116, L by V at position 120, L by I at position 120, L by T at position 120, L by Q at position 120, L by H at position 120, L by A at position 120, K by Q at position 123, K by T at position 123, K by S at position 123, K by H at position 123, R by H at position 124, R by Q at position 124, R by H at position 128, R by Q at position 128, L by V at position 130, L by I at position 130, L by T at position 130, L by Q at position 130, L by H at position 130, L by A at position 130, K by Q at position 134, K by T at position 134, K by S at position 134, K by H at position 134, K by Q at position 136, K by T at position 136, K by S at position 136, K by H at position 136, E by Q at position 137, E by H at position 137, Y by H at position 138, Y by I at position 138, R by H at position 152, R by Q at position 152, Y by H at position 155, Y by I at position 155, R by H at position 159, R by Q at position 159, Y by H at position 163, Y by I at position 163, R by H at position 165, R by Q at position 165, M by D at position 1, M by E at position 1, M by K at position 1, M by N at position 1, M by R at position 1, M by S at position 1, L by D at position 5, L by E at position 5, L by K at position 5, L by N at position 5, L by R at position 5, L by S at position 5, L by D at position 6, L by E at position 6, L by K at position 6, L by N

at position 6, L by R at position 6, L by S at position 6, L by Q at position 6, L by T at position 6, F by E at position 8, F by K at position 8, F by R at position 8, F by D at position 8, L by D at position 9, L by E at position 9, L by K at position 9, L by N at position 9, L by R at position 9, L by S at position 9, Q by D at position 10, Q by E at position 10, Q by K at position 10, Q by N at position 10, Q by R at position 10, Q by S at position 10, Q by T at position 10, S by D at position 12, S by E at position 12, S by K at position 12, S by R at position 12, S by D at position 13, S by E at position 13, S by K at position 13, S by R at position 13, S by N at position 13, S by Q at position 13, S by T at position 13, N by D at position 14, N by E at position 14, N by K at position 14, N by Q at position 14, N by R at position 14, N by S at position 14, N by T at position 14, F by D at position 15, F by E at position 15, F by K at position 15, F by R at position 15, Q by D at position 16, Q by E at position 16, Q by K at position 16, Q by N at position 16, Q by R at position 16, Q by S at position 16, Q by T at position 16, C by D at position 17, C by E at position 17, C by K at position 17, C by N at position 17, C by Q at position 17, C by R at position 17, C by S at position 17, C by T at position 17, L by N at position 20, L by Q at position 20, L by R at position 20, L by S at position 20, L by T at position 20, L by D at position 20, L by E at position 20, L by K at position 20, W by D at position 22, W by E at position 22, W by K at position 22, W by R at position 22, Q by D at position 23, Q by E at position 23, Q by K at position 23, Q by R at position 23, L by D at position 24, L by E at position 24, L by K at position 24, L by R at position 24, W by D at position 79, W by E at position 79, W by K at position 79, W by R at position 79, N by D at position 80, N by E at position 80, N by K at position 80, N by R at position 80, T by D at position 82, T by E at position 82, T by K at position 82, T by R at position 82, I by D at position 83, I by E at position 83, I by K at position 83, I by R at position 83, I by N at position 83, I by

Q at position 83, I by S at position 83, I by T at position 83, N by D at position 86, N by E at position 86, N by K at position 86, N by R at position 86, N by Q at position 86, N by S at position 86, N by T at position 86, L by D at position 87, L by E at position 87, L by K at position 87, L by R at position 87, L by N at position 87, L by Q at position 87, L by S at position 87, L by T at position 87, A by D at position 89, A by E at position 89, A by K at position 89, A by R at position 89, N by D at position 90, N by E at position 90, N by K at position 90, N by Q at position 90, N by R at position 90, N by S at position 90, N by T at position 90, V by D at position 91, V by E at position 91, V by K at position 91, V by N at position 91, V by Q at position 91, V by R at position 91, V by S at position 91, V by T at position 91, Q by D at position 94, Q by E at position 94, Q by Q at position 94, Q by N at position 94, Q by R at position 94, Q by S at position 94, Q by T at position 94, I by D at position 95, I by E at position 95, I by K at position 95, I by N at position 95, I by Q at position 95, I by R at position 95, I by S at position 95, I by T at position 95, H by D at position 97, H by E at position 97, H by K at position 97, H by N at position 97, H by Q at position 97, H by R at position 97, H by S at position 97, H by T at position 97, L by D at position 98, L by E at position 98, L by K at position 98, L by N at position 98, L by Q at position 98, L by R at position 98, L by S at position 98, L by T at position 98, V by D at position 101, V by E at position 101, V by K at position 101, V by N at position 101, V by Q at position 101, V by R at position 101, V by S at position 101, V by T at position 101, M by C at position 1, L by C at position 6, Q by C at position 10, S by C at position 13, Q by C at position 16, L by C at position 17, V by C at position 101, L by C at position 98, H by C at position 97, Q by C at position 94, V by C at position 91, N by C at position 90.

| SEQ ID NO. | Mutant |
|---|---|
| SEQ ID N° 1016 | (M1V) |

| | |
|---|---|
| SEQ ID N° 1017 | (M1I) |
| SEQ ID N° 1018 | (M1T) |
| SEQ ID N° 1019 | (M1A) |
| SEQ ID N° 1020 | (L5V) |
| SEQ ID N° 1021 | (L5I) |
| SEQ ID N° 1022 | (L5T) |
| SEQ ID N° 1023 | (L5Q) |
| SEQ ID N° 1024 | (L5H) |
| SEQ ID N° 1025 | (L5A) |
| SEQ ID N° 1026 | (F8I) |
| SEQ ID N° 1027 | (F8V) |
| SEQ ID N° 1028 | (L9V) |
| SEQ ID N° 1029 | (L9I) |
| SEQ ID N° 1030 | (L9T) |
| SEQ ID N° 1031 | (L9Q) |
| SEQ ID N° 1032 | (L9H) |
| SEQ ID N° 1033 | (L9A) |
| SEQ ID N° 1034 | (R11H) |
| SEQ ID N° 1035 | (R11Q) |
| SEQ ID N° 1036 | (F15I) |
| SEQ ID N° 1037 | (F15V) |
| SEQ ID N° 1038 | (K19Q) |
| SEQ ID N° 1039 | (K19T) |
| SEQ ID N° 1040 | (K19S) |
| SEQ ID N° 1041 | (K19H) |
| SEQ ID N° 1042 | (W22S) |
| SEQ ID N° 1043 | (W22H) |
| SEQ ID N° 1044 | (N25H) |
| SEQ ID N° 1045 | (N25S) |
| SEQ ID N° 1046 | (N25Q) |
| SEQ ID N° 1047 | (R27H) |
| SEQ ID N° 1048 | (R27Q) |

| | |
|---|---|
| SEQ ID N° 1049 | (L28V) |
| SEQ ID N° 1050 | (L28I) |
| SEQ ID N° 1051 | (L28T) |
| SEQ ID N° 1052 | (L28Q) |
| SEQ ID N° 1053 | (L28H) |
| SEQ ID N° 1054 | (L28A) |
| SEQ ID N° 1055 | (E29Q) |
| SEQ ID N° 1056 | (E29H) |
| SEQ ID N° 1057 | (Y30H) |
| SEQ ID N° 1058 | (Y30I) |
| SEQ ID N° 1059 | (L32V) |
| SEQ ID N° 1060 | (L32I) |
| SEQ ID N° 1061 | (L32T) |
| SEQ ID N° 1062 | (L32Q) |
| SEQ ID N° 1063 | (L32H) |
| SEQ ID N° 1064 | (L32A) |
| SEQ ID N° 1065 | (M1Q) |
| SEQ ID N° 1066 | (K33Q) |
| SEQ ID N° 1067 | (K33T) |
| SEQ ID N° 1068 | (K33S) |
| SEQ ID N° 1069 | (K33H) |
| SEQ ID N° 1070 | (R35H) |
| SEQ ID N° 1071 | (R35Q) |
| SEQ ID N° 1072 | (M36V) |
| SEQ ID N° 1073 | (M36I) |
| SEQ ID N° 1074 | (M36T) |
| SEQ ID N° 1075 | (M36Q) |
| SEQ ID N° 1076 | (M36A) |
| SEQ ID N° 1077 | (E85Q) |
| SEQ ID N° 1078 | (E85H) |
| SEQ ID N° 1079 | (Y92H) |
| SEQ ID N° 1080 | (Y92I) |

| | |
|---|---|
| SEQ ID N° 1081 | (K99Q) |
| SEQ ID N° 1082 | (K99T) |
| SEQ ID N° 1083 | (K99S) |
| SEQ ID N° 1084 | (K99H) |
| SEQ ID N° 1085 | (E103Q) |
| SEQ ID N° 1086 | (E103H) |
| SEQ ID N° 1087 | (E104Q) |
| SEQ ID N° 1088 | (E104H) |
| SEQ ID N° 1089 | (K105Q) |
| SEQ ID N° 1090 | (K105T) |
| SEQ ID N° 1091 | (K105S) |
| SEQ ID N° 1092 | (K105H) |
| SEQ ID N° 1093 | (Y138H) |
| SEQ ID N° 1094 | (Y138I) |
| SEQ ID N° 1095 | (R152H) |
| SEQ ID N° 1096 | (R152Q) |
| SEQ ID N° 1097 | (Y155H) |
| SEQ ID N° 1098 | (Y155I) |
| SEQ ID N° 1099 | (R159H) |
| SEQ ID N° 1100 | (R159Q) |
| SEQ ID N° 1101 | (M1D) |
| SEQ ID N° 1102 | (M1E) |
| SEQ ID N° 1103 | (M1K) |
| SEQ ID N° 1104 | (M1N) |
| SEQ ID N° 1105 | (M1R) |
| SEQ ID N° 1106 | (M1S) |
| SEQ ID N° 1107 | (L5D) |
| SEQ ID N° 1108 | (L5E) |
| SEQ ID N° 1109 | (L5K) |
| SEQ ID N° 1110 | (L5R) |
| SEQ ID N° 1111 | (L5N) |
| SEQ ID N° 1112 | (L5S) |

| | |
|---|---|
| SEQ ID N° 1113 | (L6D) |
| SEQ ID N° 1114 | (L6E) |
| SEQ ID N° 1115 | (L6K) |
| SEQ ID N° 1116 | (L6N) |
| SEQ ID N° 1117 | (L6Q) |
| SEQ ID N° 1118 | (L6R) |
| SEQ ID N° 1119 | (L6S) |
| SEQ ID N° 1120 | (L6T) |
| SEQ ID N° 1121 | (F8D) |
| SEQ ID N° 1122 | (F8E) |
| SEQ ID N° 1123 | (F8K) |
| SEQ ID N° 1124 | (F8R) |
| SEQ ID N° 1125 | (L9D) |
| SEQ ID N° 1126 | (L9E) |
| SEQ ID N° 1127 | (L9K) |
| SEQ ID N° 1128 | (L9N) |
| SEQ ID N° 1129 | (L9R) |
| SEQ ID N° 1130 | (L9S) |
| SEQ ID N° 1131 | (Q10D) |
| SEQ ID N° 1132 | (Q10E) |
| SEQ ID N° 1133 | (Q10K) |
| SEQ ID N° 1134 | (Q10N) |
| SEQ ID N° 1135 | (Q10R) |
| SEQ ID N° 1136 | (Q10S) |
| SEQ ID N° 1137 | (Q10T) |
| SEQ ID N° 1138 | (S12D) |
| SEQ ID N° 1139 | (S12E) |
| SEQ ID N° 1140 | (S12K) |
| SEQ ID N° 1141 | (S12R) |
| SEQ ID N° 1142 | (S13D) |
| SEQ ID N° 1143 | (S13E) |
| SEQ ID N° 1144 | (S13K) |

| | |
|---|---|
| SEQ ID N° 1145 | (S13N) |
| SEQ ID N° 1146 | (S13Q) |
| SEQ ID N° 1147 | (S13R) |
| SEQ ID N° 1148 | (S13T) |
| SEQ ID N° 1149 | (N14D) |
| SEQ ID N° 1150 | (N14E) |
| SEQ ID N° 1151 | (N14K) |
| SEQ ID N° 1152 | (N14Q) |
| SEQ ID N° 1153 | (N14R) |
| SEQ ID N° 1154 | (N14S) |
| SEQ ID N° 1155 | (N14T) |
| SEQ ID N° 1156 | (F15D) |
| SEQ ID N° 1157 | (F15E) |
| SEQ ID N° 1158 | (F15K) |
| SEQ ID N° 1159 | (F15R) |
| SEQ ID N° 1160 | (Q16D) |
| SEQ ID N° 1161 | (Q16E) |
| SEQ ID N° 1162 | (Q16K) |
| SEQ ID N° 1163 | (Q16N) |
| SEQ ID N° 1164 | (Q16R) |
| SEQ ID N° 1165 | (Q16S) |
| SEQ ID N° 1166 | (Q16T) |
| SEQ ID N° 1167 | (C17D) |
| SEQ ID N° 1168 | (C17E) |
| SEQ ID N° 1169 | (C17K) |
| SEQ ID N° 1170 | (C17N) |
| SEQ ID N° 1171 | (C17Q) |
| SEQ ID N° 1172 | (C17R) |
| SEQ ID N° 1173 | (C17S) |
| SEQ ID N° 1174 | (C17T) |
| SEQ ID N° 1175 | (L20N) |
| SEQ ID N° 1176 | (L20Q) |

| | |
|---|---|
| SEQ ID N° 1177 | (L20R) |
| SEQ ID N° 1178 | (L20S) |
| SEQ ID N° 1179 | (L20T) |
| SEQ ID N° 1180 | (L20D) |
| SEQ ID N° 1181 | (L20E) |
| SEQ ID N° 1182 | (L20K) |
| SEQ ID N° 1183 | (W22D) |
| SEQ ID N° 1184 | (W22E) |
| SEQ ID N° 1185 | (W22K) |
| SEQ ID N° 1186 | (W22R) |
| SEQ ID N° 1187 | (Q23D) |
| SEQ ID N° 1188 | (Q23E) |
| SEQ ID N° 1189 | (Q23K) |
| SEQ ID N° 1190 | (Q23R) |
| SEQ ID N° 1191 | (L24D) |
| SEQ ID N° 1192 | (L24E) |
| SEQ ID N° 1193 | (L24K) |
| SEQ ID N° 1194 | (L24R) |
| SEQ ID N° 1195 | (G78D) |
| SEQ ID N° 1196 | (G78E) |
| SEQ ID N° 1197 | (G78K) |
| SEQ ID N° 1198 | (G78R) |
| SEQ ID N° 1199 | (W79D) |
| SEQ ID N° 1200 | (W79E) |
| SEQ ID N° 1201 | (W79K) |
| SEQ ID N° 1202 | (W79R) |
| SEQ ID N° 1203 | (N80D) |
| SEQ ID N° 1204 | (N80E) |
| SEQ ID N° 1205 | (N80K) |
| SEQ ID N° 1206 | (N80R) |
| SEQ ID N° 1207 | (T82D) |
| SEQ ID N° 1208 | (T82E) |

| | |
|---|---|
| SEQ ID N° 1209 | (T82K) |
| SEQ ID N° 1210 | (T82R) |
| SEQ ID N° 1211 | (I83D) |
| SEQ ID N° 1212 | (I83E) |
| SEQ ID N° 1213 | (I83K) |
| SEQ ID N° 1214 | (I83R) |
| SEQ ID N° 1215 | (I83N) |
| SEQ ID N° 1216 | (I83Q) |
| SEQ ID N° 1217 | (I83S) |
| SEQ ID N° 1218 | (I83T) |
| SEQ ID N° 1219 | (N86D) |
| SEQ ID N° 1220 | (N86E) |
| SEQ ID N° 1221 | (N86K) |
| SEQ ID N° 1222 | (N86R) |
| SEQ ID N° 1223 | (N86Q) |
| SEQ ID N° 1224 | (N86S) |
| SEQ ID N° 1225 | (N86T) |
| SEQ ID N° 1226 | (L87D) |
| SEQ ID N° 1227 | (L87E) |
| SEQ ID N° 1228 | (L87K) |
| SEQ ID N° 1229 | (L87R) |
| SEQ ID N° 1230 | (L87N) |
| SEQ ID N° 1231 | (L87Q) |
| SEQ ID N° 1232 | (L87S) |
| SEQ ID N° 1233 | (L87T) |
| SEQ ID N° 1234 | (A89D) |
| SEQ ID N° 1235 | (A89E) |
| SEQ ID N° 1236 | (A89K) |
| SEQ ID N° 1237 | (A89R) |
| SEQ ID N° 1238 | (N90D) |
| SEQ ID N° 1239 | (N90E) |
| SEQ ID N° 1240 | (N90K) |

| | |
|---|---|
| SEQ ID N° 1241 | (N90Q) |
| SEQ ID N° 1242 | (N90R) |
| SEQ ID N° 1243 | (N90S) |
| SEQ ID N° 1244 | (N90T) |
| SEQ ID N° 1245 | (V91D) |
| SEQ ID N° 1246 | (V91E) |
| SEQ ID N° 1247 | (V91K) |
| SEQ ID N° 1248 | (V91N) |
| SEQ ID N° 1249 | (V91Q) |
| SEQ ID N° 1250 | (V91R) |
| SEQ ID N° 1251 | (V91S) |
| SEQ ID N° 1252 | (V91T) |
| SEQ ID N° 1253 | (Q94D) |
| SEQ ID N° 1254 | (Q94E) |
| SEQ ID N° 1255 | (Q94K) |
| SEQ ID N° 1256 | (Q94N) |
| SEQ ID N° 1257 | (Q94R) |
| SEQ ID N° 1258 | (Q94S) |
| SEQ ID N° 1259 | (Q94T) |
| SEQ ID N° 1260 | (I95D) |
| SEQ ID N° 1261 | (I95E) |
| SEQ ID N° 1262 | (I95K) |
| SEQ ID N° 1263 | (I95N) |
| SEQ ID N° 1264 | (I95Q) |
| SEQ ID N° 1265 | (I95R) |
| SEQ ID N° 1266 | (I95S) |
| SEQ ID N° 1267 | (I95T) |
| SEQ ID N° 1268 | (H97D) |
| SEQ ID N° 1269 | (H97E) |
| SEQ ID N° 1270 | (H97K) |
| SEQ ID N° 1271 | (H97N) |
| SEQ ID N° 1272 | (H97Q) |

| | |
|---|---|
| SEQ ID N° 1273 | (H97R) |
| SEQ ID N° 1274 | (H97S) |
| SEQ ID N° 1275 | (H97T) |
| SEQ ID N° 1276 | (L98D) |
| SEQ ID N° 1277 | (L98E) |
| SEQ ID N° 1278 | (L98K) |
| SEQ ID N° 1279 | (L98N) |
| SEQ ID N° 1280 | (L98Q) |
| SEQ ID N° 1281 | (L98R) |
| SEQ ID N° 1282 | (L98S) |
| SEQ ID N° 1283 | (L98T) |
| SEQ ID N° 1284 | (V101D) |
| SEQ ID N° 1285 | (V101E) |
| SEQ ID N° 1286 | (V101K) |
| SEQ ID N° 1287 | (V101N) |
| SEQ ID N° 1288 | (V101Q) |
| SEQ ID N° 1289 | (V101R) |
| SEQ ID N° 1290 | (V101S) |
| SEQ ID N° 1291 | (V101T) |
| SEQ ID N° 1292 | (M1C) |
| SEQ ID N° 1293 | (V101C) |
| SEQ ID N° 1294 | (L6C) |
| SEQ ID N° 1295 | (L98C) |
| SEQ ID N° 1296 | (Q10C) |
| SEQ ID N° 1297 | (H97C) |
| SEQ ID N° 1298 | (S13C) |
| SEQ ID N° 1299 | (Q94C) |
| SEQ ID N° 1300 | (Q16C) |
| SEQ ID N° 1301 | (N90C) |
| SEQ ID N° 1302 | (V91C) |

## I. Super-LEADs and Additiv Directional Mutagenesis (ADM)$

Also provided herein are super-LEAD mutant proteins comprising a combination of single amino acid mutations present in two or more of the respective LEAD mutant proteins. Thus, the super-LEAD mutant proteins have two of more of the single amino acid mutations derived from two or more of the respective LEAD mutant proteins. As described herein, LEAD mutant proteins provided herein are defined as mutants whose performance or fitness has been optimized with respect to the native protein. LEADs typically contain one single mutation relative to its respective native protein. This mutation represents an appropriate amino acid replacement that takes place at one is-HIT position. Further super-LEAD mutant proteins are created such that they carry on the same protein molecule, more than one LEAD mutation, each at a different is-HIT position. Once the LEAD mutant proteins have been identified using the 2D-scanning methods provided herein, super-LEADs can be generated by combining two or more individual LEAD mutant mutations using methods well-known in the art, such as recombination, mutagenesis and DNA shuffling, and by methods, such as additive directional mutagenesis and Multi-Overlapped Primer Extensions, provided herein.

### 1) Additive Directional Mutagenesis.

Also provided herein are methods for assembling on a single mutant protein multiple mutations present on the individual LEAD molecules, so as to generate super-LEAD mutant proteins. This method is referred to herein as "Additive Directional Mutagenesis" (ADM). ADM is a repetitive multi-step process where at each step after the creation of the first LEAD mutant protein a new LEAD mutation is added onto the previous LEAD mutant protein to create successive super-LEAD mutant proteins. ADM is not based on genetic recombination mechanisms, nor on shuffling methodologies; instead it is a simple one-mutation-at-a-time process, repeated as many times as necessary until the total number of desired mutations is introduced on the same molecule. To avoid the

exponentially increasing number of all possible combinations that can be generated by putting together on the same molecule a given number of single mutations, a method is provided herein that, although it does not cover all the combinatorial possible space, still captures a big part of the combinatorial potential. The word "combinatorial" is used here in its mathematical meaning (i.e., subsets of a group of elements, containing some of the elements in any possible order) and not in the molecular biological or directed evolution meaning (i.e., generating pools, or mixtures, or collections of molecules by randomly mixing their constitutive elements).

A population of sets of nucleic acid molecules encoding a collection of new super-LEAD mutant molecules is generated, tested and phenotypically characterized one-by-one in addressable arrays. super-LEAD mutant molecules are such that each molecule contains a variable number and type of LEAD mutations. Those molecules displaying further improved fitness for the particular feature being evolved, are referred to as super-LEADs. Super-LEADs may be generated by other methods known to those of skill in the art and tested by the high throughput methods herein. For purposes herein a super-LEAD typically has activity with respect to the function or biological activity of interest that differs from the improved activity of a LEAD by a desired amount, such as at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200% or more from at least one of the LEAD mutants from which it is derived. In yet other embodiments, the change in activity is at least about 2 times, 3 times, 4 times, 5 times, 6 times, 7 times, 8 times, 9 times, 10 times, 20 times, 30 times, 40 times, 50 times, 60 times, 70 times, 80 times, 90 times, 100 times, 200 times, 300 times, 400 times, 500 times, 600 times, 700 times, 800 times, 900 times, 1000 times, or more greater than at least one of the LEAD molecules from which it is derived. As with LEADs, the change in the activity for super-LEADs is dependent upon the activity that is being "evolved." The desired

alteration, which can be either an increase or a reduction in activity, will depend upon the function or property of interest.

In one embodiment provided herein, the ADM method employs a number of repetitive steps, such that at each step a new mutation is added on a given molecule. Although numerous different ways are possible for combining each LEAD mutation onto a super-LEAD protein, an exemplary way the new mutations (e.g., mutation 1 (m1), mutation 2 (m2), mutation 3 (m3), mutation 4 (m4), mutation 5 (m5), mutation n (mn)) can be added corresponds to the following diagram:

m1

m1 + m2

m1 + m2 + m3

m1 + m2 + m3 + m4

m1 + m2 + m3 + m4 + m5

m1 + m2 + m3 + m4 + m5 + ... + mn

m1 + m2 + m4

m1 + m2 + m4 + m5

m1 + m2 + m4 + m5 + ... + mn

m1 + m2 + m5

m1 + m2 + m5 + ... + mn

m2

m2 + m3

m2 + m3 + m4

m2 + m3 + m4 + m5

m2 + m3 + m4 + m5 + ... + mn

m2 + m4

m2 + m4 + m5

m2 + m4 + m5 + ... + mn

m2 + m5

m2 + m5 + ... + mn

..., etc....

**2) Multi-Overlapped Primer Extensions**

In another embodiment, provided herein is a method for the rational evolution of proteins using oligonucleotide-mediated mutagenesis referred to as "multi overlapped primer extensions." This method can be used for the rational combination of mutant LEADs to form super-LEADS. This method allows the simultaneous introduction of several mutations throughout a small protein or protein-region of known sequence. Overlapping oligonucleotides of typically around 70 bases in length (since longer oligonucleotides lead to increased error) are designed from the DNA sequence (gene) encoding the mutant LEAD proteins in such a way that they overlap with each other on a region of typically around 20 bases. These overlapping oligonucleotides (including or not point mutations) act as both template and primers in a first step of PCR (using a proofreading polymerase, e.g., Pfu DNA polymerase, to avoid unexpected mutations) to create small amounts of full-length gene. The full-length gene resulting from the first PCR is then selectively amplified in a second step of PCR using flanking primers, each one tagged with a restriction site in order to facilitate subsequent cloning. One multi overlapped extension process yields a full-length (multi-mutated) nucleic acid molecule encoding a candidate super-LEAD protein having multiple mutations therein derived from LEAD mutant proteins.

Although typically about 70 bases are used to create the overlapping oligonucleotides, the length of additional overlapping oligonucleotides for use herein can range from about 30 bases up to about 100 bases, from about 40 bases up to about 90 bases, from about 50 bases up to about 80 bases, from about 60 bases up to about 75 bases, and from about 65 bases up to about 75 bases. As set forth above, typically about 70 bases are used herein.

Likewise, although typically the overlapping region of the overlapping oligonucleotides is about 20 bases, the length of other overlapping regions for use herein can range from about 5 bases up to

about 40 bases, from about 10 bases up to about 35 bases, from about 15 bases up to about 35 bases, from about 15 bases up to about 25 bases, from about 16 bases up to about 24 bases, from about 17 bases up to about 23 bases, from about 18 bases up to about 22 bases, and from about 19 bases up to about 21 bases. As set forth above, typically about 20 bases are used herein for the overlapping region.

### J. Uses of the Mutant IFN$\alpha$, IFN$\beta$ Genes and Cytokines in Therapeutic Methods

The optimized cytokines provided herein, such as the IFN$\alpha$-2b and IFN$\beta$ proteins and other modified cytokines, are intended for use in various therapeutic as well as diagnostic methods. These include all methods for which the unmodified proteins are used. By virtue of their improved phenotypes and activities, the proteins provided herein should exhibit improvement in the corresponding *in vivo* phenotype.

In particular, the optimized cytokines, such as the IFN$\alpha$-2b and IFN$\beta$ proteins, are intended for use in therapeutic methods in which cytokines have been used for treatment. Such methods include, but are not limited to, methods of treatment of infectious diseases, allergies, microbial diseases, pregnancy related diseases, bacterial diseases, heart diseases, viral diseases, histological diseases, genetic diseases, blood related diseases, fungal diseses, adrenal diseases, cancers, liver diseases, autoimmune diseases, growth disorders, diabetes, neurodegenerative diseases, including mulitiple sclerosis, Parkinson's disease and Alzheimer's disease.

#### 1) Fusion Proteins

Fusion proteins containing a targeting agent and mutant IFN$\alpha$, including IFN$\alpha$-2b and IFN$\alpha$-2a, and IFN$\beta$ mutant proteins, or cytokine protein also are provided. Pharmaceutical compositions containing such fusion proteins formulated for administration by a suitable route are provided. Fusion proteins are formed by linking in any order the mutant protein and an agent, such as an antibody or fragment thereof, growth

factor, receptor, ligand and other such agent for directing the mutant protein to a targeted cell or tissue. Linkage can be effected directly or indirectly via a linker. The fusion proteins can be produced recombinantly or chemically by chemical linkage, such as via heterobifunctional agents or thiol linkages or other such linkages. The fusion proteins can contain additional components, such as *E. coli* maltose binding protein (MBP) that aid in uptake of the protein by cells (see, International PCT application No. WO 01/32711).

**2) Nucleic Acid Molecules for Expression**

Nucleic acid molecules encoding the mutant cytokines including the mutant IFN$\beta$ proteins and IFN $\alpha$ proteins, such as the IFN$\alpha$-2b and IFN$\alpha$-2a proteins, provided herein, or the fusion protein operably linked to a promoter, such as an inducible promoter for expression in mammalian cells also are provided. Such promoters include, but are not limited to, CMV and SV40 promoters; adenovirus promoters, such as the E2 gene promoter, which is responsive to the HPV E7 oncoprotein; a PV promoter, such as the PBV p89 promoter that is responsive to the PV E2 protein; and other promoters that are activated by the HIV or PV or oncogenes.

The mutant cytokines inlcuding the mutant interferons (IFN$\alpha$'s and IFN$\beta$') proteins provided herein, also can be delivered to the cells in gene transfer vectors. The transfer vectors also can encode encode additional other therapeutic agent(s) for treatment of the disease or disorder, such cancer or HIV infection, for which the cytokine is administered.

**3) Formulation of Optimized Cytokines and methods of treatment**

Pharmaceutical compositions containing an optimized cytokine produced herein, such as IFN$\alpha$-2b, IFN$\alpha$-2a and IFN$\beta$, fusion proteins or encoding nucleic acid molecules can be formulated in any conventional manner by mixing a selected amount of an optimized cytokine with one or more physiologically acceptable carriers or excipients. Selection of the carrier or excipient depends upon the mode of administration (*i.e.*,

systemic, local, topical or any other mode) and disorder treated. The pharmaceutical compositions provided herein can be formulated for single dosage administration. The concentrations of the compounds in the formulations are effective for delivery of an amount, upon administration, that is effective for the intended treatment. Typically, the compositions are formulated for single dosage administration. To formulate a composition, the weight fraction of a compound or mixture thereof is dissolved, suspended, dispersed or otherwise mixed in a selected vehicle at an effective concentration such that the treated condition is relieved or ameliorated. Pharmaceutical carriers or vehicles suitable for administration of the compounds provided herein include any such carriers known to those skilled in the art to be suitable for the particular mode of administration.

In addition, the compounds may be formulated as the sole pharmaceutically active ingredient in the composition or may be combined with other active ingredients. Liposomal suspensions, including tissue-targeted liposomes, may also be suitable as pharmaceutically acceptable carriers. These may be prepared according to methods known to those skilled in the art. For example, liposome formulations may be prepared as described in U.S. Patent No. 4,522,811.

The active compound is included in the pharmaceutically acceptable carrier in an amount sufficient to exert a therapeutically useful effect in the absence of undesirable side effects on the patient treated. The therapeutically effective concentration may be determined empirically by testing the compounds in known *in vitro* and *in vivo* systems, such as the assays provided herein. The active compounds can be administered by any appropriate route, for example, orally, parenterally, intravenously, intradermally, subcutaneously, or topically, in liquid, semi-liquid or solid form and are formulated in a manner suitable for each route of administration.

The optimized cytokine and physiologically acceptable salts and solvates can be formulated for administration by inhalation (either through the mouth or the nose) or for oral, buccal, parenteral or rectal administration. For administration by inhalation, the optimized cytokine can be delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluorethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g., gelatin for use in an inhaler or insufflator can be formulated containing a powder mix of a therapeutic compound and a suitable powder base such as lactose or starch.

For oral administration, the pharmaceutical compositions can take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets can be coated by methods well known in the art. Liquid preparations for oral administration can take the form of, for example, solutions, syrups or suspensions, or they can be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations can be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also

contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration can be suitably formulated to give controlled release of the active compound. For buccal administration the compositions can take the form of tablets or lozenges formulated in conventional manner.

The optimized cytokine can be formulated for parenteral administration by injection e.g., by bolus injection or continuous infusion. Formulations for injection can be presented in unit dosage form e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions can take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder-lyophilized form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

In addition to the formulations described previously, the optimized cytokine also can be formulated as a depot preparation. Such long acting formulations can be administered by implantation (for example, subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the therapeutic compounds can be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The active agents can be formulated for local or topical application, such as for topical application to the skin and mucous membranes, such as in the eye, in the form of gels, creams, and lotions and for application to the eye or for intracisternal or intraspinal application. Such solutions, particularly those intended for ophthalmic use, can be formulated as 0.01% - 10% isotonic solutions, pH about 5-7, with appropriate salts. The compounds can be formulated as aerosols for topical application, such as by inhalation (see, *e.g.*, U.S. Patent Nos. 4,044,126, 4,414,209,

and 4,364,923, which describe aerosols for delivery of a steroid useful for treatment inflammatory diseases, particularly asthma).

The concentration of active compound in the drug composition will depend on absorption, inactivation and excretion rates of the active compound, the dosage schedule, and amount administered as well as other factors known to those of skill in the art. For example, the amount that is delivered is sufficient to treat the symptoms of hypertension.

The compositions, if desired, can be presented in a package, in kit a or dispenser device, that can contain one or more unit dosage forms containing the active ingredient. The package, for example, contains metal or plastic foil, such as a blister pack. The pack or dispenser device can be accompanied by instructions for administration. The compositions containing the active agents can be packaged as articles of manufacture containing packaging material, an agent provided herein, and a label that indicates the disorder for which the agent is provided.

Methods of treatment of cytokine-mediated or cytokine-involved diseases and immunotherapeutic methods are provided. The modified cytokines can be used in any method of treatment for which the unmodified cytokine is used. Hence the modified cytokines can be used for treatment of all disorders noted herein for the respective cytokines and for those known to those of skill in the art for each of the others, such as immunotherapeutic treatment (interleukins) and red blood cell expansion and stem cell expansion. The following table summarizes exemplary uses in addition to those noted herein of exemplary modified cytokines provided herein:

| Cytokine | Exemplary Uses, Diseases and Treatment |
|---|---|
| IL-10 | anti-inflammatory treatment of chronic liver injury and disease; myeloma |
| Interferon-gamma | interstitial/idiopathic pulmonary fibrosis; adjunctive immunotherapy for immunosupressed patients |
| Granulocyte colony stimulating factor | Crohn's disease; cardiac disease; acquired and congenital neutropenias; asthma |

| Cytokine | Exemplary Uses, Dis ases and Tr atment |
|---|---|
| Leukemia inhibitory factor | myocardial infarction; multiple sclerosis; prevention of axonal atrophy;olfactory epithelium replacement stimulation |
| Human growth hormone | growth hormone deficiency; acromegaly |
| Ciliary neurotrophic factor | retinal degeneration treatments; neurodegnerative diseases such as Huntingtons; auditory degenerative diseases |
| Leptin | obesity; pancreatitis; endometreosis |
| Oncostatin M | chronic infammatory diseases; rheumatoid arthritis; multiple sclerosis; tissue damage supression |
| Interleukin-6 | protection from liver injury; Crohn's disease; hematopoietic associated diseases |
| Interleukin-12 | coksakievirus treatment;neuroblastoma; melanoma, renal cell carcinoma; mucosal immunity induction |
| Erythropoietin | hypoxia; myocardial ischemia; anemia with renal failure and cancer treatments |
| Granulocyte-macrophage colony stimulating factor | stimulate antigen presenting cells; anti-tumor activity for leukemia, melanoma, and breast, liver and renal cell carcinomas; adjunctive immunotherapy for immunosupressed patients; automimmune disease |
| Interleukin-2 | immune reactivation after chemotherapy; melanoma; colon carcinoma |
| Interleukin-3 | leukemia cell targeting; motor neuropathy; amyotrophic lateral sclerosis; asthma |
| Interleukin-4 | allergic asthma; lupus |
| Interleukin-5 | treatment for parasites;asthma; allergic diseases accompanied by eosinophilia |
| Interleukin-13 | intracellular infections; B-cell cancers; asthma |
| Flt3 ligand | prostatate cancer; myeloid leukemia; engraftment of allogenic hematopietic stem cells |
| Stem cell factor | hepatic injury; asthma; hematopoietic engraftment |

Treatment can be effected by any suitable route of administration using suitable formulations. If necessary, a particular dosage and duration and treatment protocol can be empirically determined or extrapolated.

**K. Examples**

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention. The specific methods exemplified can be practiced with other species. The examples are intended to exemplify generic processes.

**EXAMPLE 1**

This example describes a plurality of chronological steps including steps from (i) to (viii):

(i) cloning of IFN$\alpha$ cDNA in a mammalian cell expression plasmid (section A.1)

(ii) generation of a collection of targeted mutants on the IFN$\alpha$ cDNA in the mammalian cell expression plasmid (section B)

(iii) production of IFN$\alpha$ mutants in mammalian cells (section C.1)

(iv) screening and partial in vitro characterization of IFN$\alpha$ mutants produced in mammalian cells in search of lead mutants (section D)

(v) cloning of the lead mutants into a bacterial cell expression plasmid (section A.2)

(vi) expression of lead mutants in bacterial cells (section C.2)

(vii) *in vitro* characterization of lead mutants produced in bacteria (section D)

(viii) in vivo characterization of lead mutants produced in bacteria (section E).

**A. Cloning of IFN$\alpha$-2b encoding cDNA**

**A.1. Cloning of IFN $\alpha$-2b cDNA in a mammalian cell expression plasmid**

The IFN $\alpha$-2b cDNA was first cloned into an mammalian expression vector, prior to the generation of the selected mutations. A collection of

mutants was then generated such that each individual mutant was created and processed individually, physically separated form each other and in addressable arrays. The mammalian expression vector pSSV9 CMV 0.3 pA was engineered as follows:

The pSSV9 CMV 0.3 pA was cut by *Pvu*II and religated (this step gets rid of the ITR functions), prior to the introduction of a new *Eco*RI restriction site by Quickchange mutagenesis (Stratagene). The oligonucleotides primers were:

EcoRI forward primer 5'-GCCTGTATGATTTATTGGATGTTGGAATTCC-CTGATGCGGTATTTTCTCCTTACG-3' (SEQ ID NO: 218)

EcoRI reverse primer 5'-CGTAAGGAGAAAATACCGCATCAGGGAATT-CCAACATCCAATAAATCATACAGGC-3' (SEQ ID NO: 219).

The construct sequence was confirmed by using the following oligonucleotides:

Seq ClaI forward primer: 5'-CTGATTATCAACCGGGGTACATATGATTGAC-ATGC-3' (SEQ ID NO: 220)

Seq XmnI reverse primer 5'-TACGGGATAATACCGCGCCACATAGCAGAA-C-3' (SEQ ID NO: 221).

Then, the *Xmn*I-*Cla*I fragment containing the newly introduced *Eco*RI site was cloned into pSSV9 CMV 0.3 pA (SSV9 is a clone containing the entire adeno-associated virus (AAV) genome inserted into the PvuII site of plasmid pEMBL (see, Du *et al.* (1996) *Gene Ther* *3*:254-261)) to replace the corresponding wild-type fragment and produce construct pSSV9-2EcoRI.

The DNA sequence of the IFN$\alpha$-2b cDNA, which was inserted into the mammalian vector pDG6 (ATCC accession No. 53169), was confirmed using a pair of internal primers. The sequences of the IFN$\alpha$-2b-related oligonucleotides for sequencing follow:

Seq forward primer: 5'-CCTGATGAAGGAGGACTC-3' (SEQ ID NO: 222)

Seq reverse primer: 5'-CCAAGCAGCAGATGAGTC-3' (SEQ ID NO: 223).

Since the beginning of the IFN$\alpha$-2b encoding cDNA (the signal peptide encoding sequence) is absent in pDG6, it was added using the oligonucleotide (see below)to the amplified gene. First, the IFN$\alpha$-2b cDNA was amplified by PCR using pDG6 as template using the following oligonucleotides as primers:

IFN$\alpha$-2b 5' primer 5'-TCAGCTGCAAGTCAAGCTGCTCTGTGGGCTG-3' (SEQ ID NO: 224)

IFN$\alpha$-2b 3' primer 5'-GCTCTAGATCATTCCTTACTTCTTAAACTTTC-TTGCAAGTTTGTTGAC-3' (SEQ ID NO: 225)

The PCR product was then used in an overlapping PCR using the following oligonucleotide sequences, having *Hind* III or *Xba*I restriction sites (underlined) or the DNA sequence missing in pDG6 (underlined):

IFN$\alpha$-2b HindIII primer 5'-CCCAAGCTTATGGCCTTGACCTTTGCTTTACT-GGTG-3' (SEQ ID NO: 226)

IFN$\alpha$-2b XbaI primer 5'-GCTCTAGATCATTCCTTACTTCTTAAACTTTC-TTGCAAGTTTGTTGAC-3' (SEQ ID NO: 227)

IFN$\alpha$-2b 80bp 5' primer 5'-CCCAAGCTT**ATGGCCTTGACCTTTGCTTTA-CTGGTGGCCCTCCTGGTGC**TCAGCTGCAAGTCAAGCTGCTCTGTGGGCTG-3' (SEQ ID NO: 228).

The entire IFN$\alpha$-2b cDNA was cloned into the pTOPO-TA vector (Invitrogen). After checking gene sequence by automatic DNA sequencing, the *Hin*dIII-*Xba*I fragment containing the gene of interest was subcloned into the corresponding sites of pSSV9-2EcoRI to produce pAAV-EcoRI-IFNalpha-2b (pNB-AAV-IFN alpha-2b).

### A.2 Cloning of the IFN $\alpha$-2b leads in an E. coli expression plasmid

#### A.2.1 Characterization of the bacterial cells

BL21-CodonPlus(DE3)-RP® competent Escherichia coli cells are derived from Stratagene's high-performance BL21-Gold competent cells. These cells enable efficient high-level expression of heterologous proteins in *E. coli*. Efficient production of heterologous proteins in E. coli is frequently limited by the rarity, in E.coli, of certain tRNAs that are abundant in the organisms from which the heterologous proteins are derived. Availability of tRNAs allows high-level expression of many

heterologous recombinant genes in BL21-Codon Plus cells that are poorly expressed in conventional BL21 strains.BL21-CodonPLus(DE3)-RP cells contain a ColE1-compatible, pACYC-based plasmid containing extra copies of the argU and proL tRNA genes.

**A.2.2 Cloning of wild-type IFN *α***

To express IFN *α*-2b in *E.coli* cDNA encoding the mature form of IFN-2 *α*-2b was finally cloned into the plasmid pET-11 (Novagen). Briefly, this cDNA fragment was amplified by PCR using the primers SEQ ID Nos. 1306 and 1305, respectively:

FOR-IFNA-5' AACATATGTGTGATCTGCCTCAAACCCACAGCCTGGGTAGC 3'

REV-IFNA-5' AAGGATCCTCATTCCTTACTTCTTAAACTTTCTTGCAAGTTTGTTG3',

from pSSV9-EcoRI-IFN *α*-2b (see above), which contains full-length IFN-2 alpha cDNA as a matrix, using Herculase DNA-polymerase (Stratagene). The PCR fragment was subcloned into pTOPO-TA vector (Invitrogen) yielding pTOPO-IFN *α*-2b. The sequence was verified by sequencing. pET11 IFN *α*-2b was prepared by insertion of the *NdeI-Bam HI* (Biolabs) fragment from pTOPO-IFN *α*-2b into the *NdeI-Bam HI* sites of pET 11. The DNA sequence of the resulting pET 11-IFN *α*-2b construct was verified by sequencing and the plasmid was used for IFN *α*-2b expression in *E.coli*.

**A.2.3 Cloning of IFN *α*-2b mutants from the mammalian expression plasmid into the E.coli expression plasmid**

Lead mutants of Interferon alpha were first generated in the pSSV9-IFNa-EcoRI plasmid. With the only exception of E159H and E159Q, all mutants were amplified using the primers below. Primers contained NdeI (in Forward) and BamHI (in Reverse) restriction sites:

FOR-IFNA-5' AACATATGTGTGATCTGCCTCAAACCCACAGCCTGGGTAGC 3' (SEQ ID No. 1306; and

REV-IFNA-5' AAGGATCCTCATTCCTTACTTCTTAAACTTTCTTGCAAGTTTGTTG 3' (SEQ ID No. 1305)

Mutants E159H and E159Q were amplified using the following primers on reverse side (primer forward was the same than described above):

REV-IFNA-E159H-5' AAGGATCCTCATTCCTTACTTCTTAAACTGTGTTGCAAGTTTGTTG 3' SEQ ID No. 1304 above; and

REV-IFNA-E159Q-5' AAGGATCCTCATTCCTTACTTCTTAAACTCTGTTGCAAGTTTGTTG 3' SEQ ID No. 1305.

Mutants were amplified with Pfu Turbo Polymerase (Stratagene) according. PCR products were cloned into pTOPO plasmid (Zero Blunt TOPO PCR cloning kit, Invitrogen). The presence of the desired mutations was checked by automatic sequencing. The NdeI + BamHI fragment of the pTOPO-IFNa positive clones was then cloned into NdeI + BamHI sites of the pET11 plasmid.

**B. Construction of a collection of IFN$\alpha$-2b mutants in a mammalian expression plasmid**

A series of mutagenic primers was designed to generate the appropriate site-specific mutations in the IFN$\alpha$-2b cDNA. Mutagenesis reactions were performed with the Chameleon® mutagenesis kit (Stratagene) using pNB-AAV-IFN$\alpha$-2b as the template. Each individual mutagenesis reaction was designed to generate one single mutant protein. Each individual mutagenesis reaction contains one and only one mutagenic primer. For each reaction, 25 pmoles of each (phosphorylated) mutagenic primer were mixed with 0.25 pmoles of template, 25 pmoles of selection primer (introducing a new restriction site), and 2 $\mu$l of 10X mutagenesis buffer (100 mM Tris-acetate pH 7.5; 100 mM MgOAc; 500 mM KOAc pH 7.5) into each well of 96 well-plates. To allow DNA annealing, PCR plates were incubated at 98 °C during 5 min and immediately placed 5 min on ice, before incubating at room temperature during 30 min. Elongation and ligation reactions were allowed by addition of 7 $\mu$l of nucleotide mix (2.86 mM each nucleotide; 1.43 X mutagenesis buffer) and 3 $\mu$l of a freshly prepared enzyme mixture of dilution buffer (20 mM Tris HCl pH7.5; 10 mM KCl; 10 mM $\beta$-mercaptoethanol; 1 mM DTT; 0.1 mM EDTA; 50 % glycerol), native T7 DNA polymerase (0.025 U/$\mu$l), and T4 DNA ligase (1 U/$\mu$l) in a ratio of 1:10, respectively. Reactions were incubated at 37 °C for 1 h before inactivation of T4 DNA

ligase at 72 °C during 15 min. In order to eliminate the parental plasmid, 30 µl of a mixture containing 1X enzyme buffer and 10 U of restriction enzyme was added to the mutagenic reactions followed by incubation at 37 °C for at least 3 hours. Next, 90 µl aliquots of XL*mutS* competent cells (Stratagene) containing 25 mM β-mercaptoethanol were place in ice-chilled deep-well plates. Then, plates were incubated on ice for 10 min with gentle vortex every 2 min. Transformation of competent cells was performed by adding aliquots of the restriction reactions (1/10 of reaction volume) and incubating on ice for 30 min. A heat pulse was performed in a 42 °C water bath for 45 s, followed by incubation on ice for 2 minutes. Preheated SOC medium (0.45 ml) was added to each well and plates were incubated at 37 °C for 1 h with shaking. In order to enrich for mutated plasmids, 1 ml of 2 X YT broth medium supplemented with 100 µg/ml ampicillin was added to each transformation mixture followed by overnight incubation at 37 °C with shaking. Plasmid DNA isolation was performed by alkaline lysis using Nucleospin Multi-96 Plus Plasmid Kit (Macherey-Nagel) according to the manufacturer's instructions. Selection of mutated plasmids was performed by digesting 500 µg of plasmid preparation with 10 U of selection endonuclease in an overnight incubation at 37 °C. A fraction of the digested reactions (1/10 of the total volume) was transformed into 40 µl of Epicurian coli XL1-Blue competent cells (Stratagene) supplemented with 25 mM β-mercaptoethanol.

Transformation was performed was as described above. Transformants were selected on LB-ampicillin agar plates incubated overnight at 37 °C. Isolated colonies were picked up and grown overnight at 37 °C into deep-well plates. Four clones per reaction were screened by endonuclease digestion of a new restriction site introduced by the selection primer. Finally, each mutation that was introduced to produce this collection of candidate LEAD IFNα-2b mutant plasmids

encoding the proteins set forth in Table 2 of Example 2 was confirmed by automatic DNA sequencing.

**C. Production of IFNa-2b mutants**

**C.1 In mammalian cells**

IFN$\alpha$-2b mutants were produced in 293 human embryo kidney (HEK) cells (obtained from ATCC), using Dubelcco's modified Eagle's medium supplemented with glucose (4.5 g/L; Gibco-BRL) and fetal bovine serum (10%, Hyclone). Cells were transiently transfected with the plasmids encoding the IFN$\alpha$-2b mutants as follows: 0.6 x $10^5$ cells were seeded into 6 well-plates and grown for 36 h before transfection Confluent cells at about 70%, were supplemented with 2.5 $\mu$g of plasmid (IFN$\alpha$-2b mutants) and 10 mM poly-ethylene-imine (25 KDa PEI, Sigma-Aldrich). After gently shaking, cells were incubated for 16 h. Then, the culture medium was changed with 1 ml of fresh medium supplemented with 1% of serum. IFN$\alpha$-2b was measured on culture supernatants obtained 40 h after transfection and stored in aliquots at -80 °C until use.

Supernatants containing IFN$\alpha$-2b from transfected cells were screened following sequential biological assays as follows. Normalization of IFN$\alpha$-2b concentration from culture supernatants was performed by enzyme-linked immunoabsorbent assay (ELISA) using a commercial kit (R & D) and following the manufacturer's instructions. This assay includes plates coated with an IFN$\alpha$-2b monoclonal antibody that can be developed by coupling a secondary antibody conjugated to the horseradish peroxidase (HRP). IFN$\alpha$-2b concentrations on samples containing (i) wild type IFN$\alpha$-2b produced under comparable conditions as the mutants, (ii) the IFN$\alpha$-2b mutants and (iii) control samples(produced from cells expressing GFP) were estimated by using an international reference standard provided by the NIBSC, UK.

**C.2 In bacteria**

A volume of 200 ml of culture medium (LB/Ampicillin/ Chloramphenicol) was inoculated with 5 ml of pre-culture BL21-

pCodon+-pET-IFN $\alpha$-2b muta overnight at 37 °C with constant shaking (225 rpm). The production of IFN $\alpha$-2b was induced by the addition of 50 $\mu$l of 2M IPTG at $DO_{600nm} \sim 0.6$.

The culture was continued for 3 additional hours and was centrifuged at 4°C and 5000 g for 15 minutes. The supernatant (culture medium) was discarded and bacteria were lysed in 8 ml of lysis buffer by thermal shock (freezing – thawing: 37°C – 15 min; -80°C – 10 min; 37°C – 15 min; -80°C – 10 min; 37°C – 15 min). After centrifugation (10000 g, 15 min, 4°C), the supernatant (soluble proteins fraction) was discarded, and the precipitated material (insoluble protein fraction containing the IFN $\alpha$ -2b protein as inclusion bodies) was purified.

### C.3 Pre-purification of IFN $\alpha$-2b as inclusion bodies in *E. coli*

#### C.3.1 Washing of inclusion bodies by sonication

Pellets containing the inclusion bodies were suspended in 10 ml of buffer and sonicated (80 watts) on ice, 1 second "on," 1 second "off" for a total of 4 min. Suspensions were then centrifuged (4°C, 10000 g, 15 min), and supernatants were recovered. Pellets were resuspended in 10 ml of buffer for a new sonication/centrifugation cycle. Triton X-100 was then eliminated by two additional cycles of sonication/centrifugation with buffer. Pellets containing the inclusion bodies were recovered and dissolved. The washed supernatants were stored at 4°C.

#### C.3.2 Solubilization of inclusion bodies by denaturation

Once washed, the inclusion bodies were solubilized in buffer at a concentration estimated in 0.3 mg/ml measuring the $OD_{280}$ (considering the coefficient of molar extinction of IFN $\alpha$-2b). Solubilization was carried out overnight at 4°C, under shaking.

#### C.3.3 Renaturation of IFN $\alpha$-2b by dialysis of GdnHCl

Samples contained 1 mg of protein at 0.3 mg/ml (5 ml in total) in buffer. The GdnHCl (Hydrochloride Guanidium) present in the samples was eliminated by dialysis (minimum membrane cut = 10 kDa) overnight at 4°C against buffer (1litre) (final concentration of GdnHCl : 43 Mm).

Next, samples were further dialysed against 1litre of buffer during 2:30h. This step was repeated two additional times. After dialysis, very little precipitate was visible.

**D. Screening and *in vitro* charaterization of IFN α-2b mutants**

Two activities were measured directly on IFN samples: antiviral and antiproliferation activities. Dose (concentration) - response (activity) experiments for antiviral or antiproliferation activity permitted calculation of the 'potency' for antiviral and antiproliferation activities, respectively. Antiviral and antiproliferation activities also were measured after incubation with proteolytic samples, such as specific proteases, mixtures of selected proteases, human serum or human blood. Assessment of activity following incubation with proteolytic samples allowed to determine the residual (antiviral or antiproliferation) activity and the respective kinetics of half-life upon exposure to proteases.

**D.1. Antiviral activity**

IFNα-2b protects cells against viral infection by a complex mechanism devoted to create an unfavorable environment for viral proliferation. Cellular antiviral response due to IFNα-2b (IFN anti-viral assay) was assessed using an interferon-sensitive HeLa cell line (ATCC accession no. CCL-2) treated with the encephalomyocarditis virus (EMCV). The assessment of either the virus-induced cytopathic effects (CPE) or the amount of EMCV mRNA in extracts of infected cells by RT-PCR was used to determine IFNα activity in samples.

**D.1.1 Antiviral activity - measure by RT-qPCR**

Confluent cells were trypsinized and plated at density 2 x $10^4$ cells/well in DMEM 5% SVF medium (Day 0). Cells were incubated with IFN α-2b (at a concentration of 500 U/ml) to get 500 pg/ml and 150 pg/well (100 µl of IFN solution), during 24 h at 37 °C prior to be challenged with EMCV (1/1000 dilution; MOI 100). After an incubation of 16 h, when virus–induced CPE was near maximum in untreated cells, the number of EMCV particles in each well was determined by RT-PCR

quantification of EMCV mRNA, using lysates of infected cells. RNA from cell extracts was purified after a DNAse/proteinase K treatment (Applied Biosystems). The CPE was evaluated using both Uptibleu (Interchim) and MTS (Promega) methods, which are based on detecting bio-reductions produced by the metabolic activity of cells in a flourometric and colorimetric manner, respectively. In order to produce a standard curve for EMCV quantification, a 22 bp DNA fragment of the capsid protein-cDNA was amplified by PCR and cloned into pTOPO-TA vector (Invitrogen). Next, RT-PCR quantification of known amounts of pTOPO-TA-EMCV capsid gene was performed using the One-step RT-PCR kit (Applied Biosystems) and the following EMCV-related (cloning) oligonucleotides and probe:

EMCV forward primer 5'-CCCCTACATTGAGGCATCCA-3' (SEQ ID NO: 229)

EMCV reverse primer 5'-CAGGAGCAGGACAAGGTCACT-3' (SEQ ID NO: 230)

EMCV probe 5'-(FAM)CAGCCGTCAAGACCCAACCGCT(TAMR A)-3' (SEQ ID NO: 231).

**D.1.2 Antiviral activity - measure by CPE**

Antiviral activity of IFN $\alpha$-2b was determined by the capacity of the cytokine to protect Hela cells against EMC (mouse encephalomyocarditis) virus-induced cytopathic effects. The day before, Hela cells ($2x10^5$ cells/ml) were seeded in flat-bottomed 96-well plates containing 100 $\mu$l/well of Dulbecco's MEM-Glutamaxl-sodium pyruvate medium supplemented with 5% SVF and 0.2% of gentamicin. Cells were growth at 37°C in an atmosphere of 5% $CO_2$ for 24 hours.

Two-fold serial dilutions of interferon samples were made with MEM complete media into 96-Deep-Well plates with final concentration ranging from 1600 to 0.6 pg/ml. The medium was aspirated from each well and 100 $\mu$l of interferon dilutions were added to Hela cells. Each interferon sample dilution was assessed in triplicate. The two last rows of the plates were filled with 100 $\mu$l of medium without interferon dilution samples in order to serve as controls for cells with and without virus.

After 24 hours of growth, a 1/1000 EMC virus dilution solution was placed in each well except for the cell control row. Plates were returned to the $CO_2$ incubator for 48 hours. Then, the medium was aspirated and the cells were stained for 1 hour with 100 $\mu$l of Blue staining solutio to determine the proportion of intact cells. Plates were washed in a distilled water bath. The cell bound dye was extracted using 100 $\mu$l of ethylene-glycol mono-ethyl-ether (Sigma). The absorbance of the dye was measured using an Elisa plate reader (Spectramax). The antiviral activity of IFN $\alpha$-2b samples (expressed as number of IU/mg of proteins) was determined as the concentration needed for 50% protection of the cells against EMC virus-induced cytopathic effects. For proteolysis experiments, each point of for the kinetic measurements was assessed at 500 and 166 pg/ml in triplicate.

**D.2 Antiproliferation activity**

Anti-proliferative activity of interferon$\alpha$-2b was determined by the capacity of the cytokine to inhibit proliferation of Daudi cells. Daudi cells ($1x10^4$ cells) were seeded in flat-bottomed 96-well plates containing 50$\mu$l/well of RPMI 1640 medium supplemented with 10% SVF, 1X glutamin and 1ml of gentamicin. No cell was added to the last row ("H" row) of the flat-bottomed 96-well plates in order to evaluate background absorbance of culture medium.

At the same time, two-fold serial dilutions of interferon samples were made with RPMI 1640 complete media into 96-Deep-Well plates with final concentration ranging from 6000 to 2.9 pg/ml. Interferon dilutions (50$\mu$l) were added to each well containing 50$\mu$l of Daudi cells. The total volume in each well should now be 100$\mu$l. Each interferon sample dilution was assessed in triplicate. Each well of the "G" row of the plates was filled with 50$\mu$l of RPMI 1640 complete media in order to be used as positive control. The plates are incubated for 72 hours at 37°C in a humidified, 5% CO2 atmosphere.

After 72 hours of growth, 20 µl of Cell titer 96 Aqueous one solution reagent (Promega) was added to each well and incubated 1H30 at 37°C in an atmosphere of 5% $CO_2$. To measure the amount of colored soluble formazan produced by cellular reduction of the MTS, the absorbance of the dye was measured using an Elisa plate reader (spectramax) at 490nm.

The corrected absorbances ("H" row background value subtracted) obtained at 490nm were plotted versus concentration of cytokine. The ED50 value was calculated by determining the X-axis value corresponding to one-half the difference between the maximum and minimum absorbance values. (ED50 = the concentration of cytokine necessary to give one-half the maximum response).

**D.3 Treatment of IFN α-2b with proteolytic preparations**

Mutants were treated with proteases in order to identify resistant molecules. The resistance of the mutant IFN α-2b molecules compared to wild-type IFN α-2b against enzymatic cleavage (30 min, 25 °C) by a mixture of proteases (containing 1.5 pg of each of the following proteases (1% wt/wt, Sigma): α-chymotrypsin, carboxypeptidase, endoproteinase Arg-C, endoproteinase Asp-N, endoproteinase Glu-C, endoproteinase Lys-C, and trypsin) was determined. At the end of the incubation time, 10 µl of anti-proteases complete, mini EDTA free, Roche (one tablet was dissolved in 10 ml of DMEM and then diluted to 1/1000) was added to each reaction in order to inhibit protease activity. Treated samples were then used to determine residual antiviral or antiproliferation activities.

**D.4 Protease resistance - Kinetic analysis**

The percent of residual IFN α-2b activity over time of exposure to proteases was evaluated by a kinetic study using either (a) 15 pg of chymotrypsin (10%wt/wt), (b) a lysate of human blood at dilution 1/100, (c) 1.5 pg of protease mixture, or (d) human serum. Incubation times were: 0 h, 0.5 h, 1 h, 4 h, 8 h, 16 h, 24 h and 48 h. Briefly, 20 µl of

each proteolytic sample (proteases, serum, bnlood) was added to 100 µl of IFN α-2b at 1500 pg/ml (500U/ml) and incubated for variable times, as indicated. At the appropriate time points, 10 µl of anti-proteases mixture, mini EDTA free, Roche (one tablet was dissolved in 10 ml of DMEM and then diluted to 1/500) was added to each well in order to stop proteolysis reactions. Biological activity assays were then performed as described for each sample in order to determine the residual activity at each time point.

**D.5 Performance**

The various biological activities, protease resistance and potency of each individual mutant were analyzed using a mathematical model and algorithm (NautScan™; described in French Patent No. 9915884; (published as International PCT application No. WO 01/44809 based on PCT n° PCT/FR00/03503). Data was processed using a Hill equation-based model that uses key feature indicators of the performance of each individual mutant. Mutants were ranked based on the values of their individual performance and those on the top of the ranking list were selected as leads.

**E. Pharmacokinetics of selected lead mutants in mice**

IFNα –2b mutants selected on the basis of their overall performance in vitro, were tested for pharmacokinetics in mice in order to have an indication of their half-life in blood in vivo. Mice were treated by subcutaneous (SC) injection with alicuots of each of a number of selected lead mutants. Blood was collected at increasing time points between 0.5 and 48 hs after injection. Inmediatedly after collection, 20 ml of anti-protease solution were added to each blood sample. Serum was obtained for further analysis.Residual IFN-α activity in blood was determined using the tests described in the precedent sections for in vitro characterization. Wild-type IFN α (that had been produced in bacteria under comparable conditions as the lead mutants) as well as a pegylated derivative of IFN α, Pegasys (Roche), also were tested for pharmacokinetics in the same experiments.

## EXAMPLE 2

This example demonstrates the 2-dimensional (2D)scanning of IFN$\alpha$-2b for increased resistance to proteolysis. For results, see Figures 6(A)-6(N), 6(T) and 6(U).

**A) Identifying some or all possible target sites on the protein sequence that are susceptible to digestion by one or more specific proteases (these sites are the is-HITs).**

Because IFN$\alpha$-2b is administered as a therapeutic protein in the blood stream, a set of proteases was identified that were expected to broadly mimic the protease contents in serum. From that list of proteases, a list of the corresponding target amino acids was identified (shown in parenthesis) as follows: $\alpha$-chymotrypsin (F, L, M, W, and Y), endoproteinase Arg-C (R), endoproteinase Asp-N (D), endoproteinase Glu-C (E), endoproteinase Lys-C (K), and trypsin (K and R) Carboxypeptidase Y, which cleaves non-specifically from the carboxy-terminal ends of proteins, was also included in the protease mixture. The distribution of the target amino acids over the protein sequence spreads over the complete length of the protein, suggesting that the protein is potentially sensitive to protease digestion all over its sequence (FIG6A). In order to restrict the number of is-HITs to a lower number of candidate positions, the 3-dimensional structure of the IFN$\alpha$-2b molecule (PDB code 1RH2) was used to identify and select only those residues exposed on the surface, while discarding from the candidate list those which remain buried in the structure, and therefore stay less susceptible to proteolysis (FIG6B).

**B) Identifying appropriate replacing amino acids, specific for each is-HIT, such that if used to replace one or more of the original, such as native, amino acids at that specific is-HIT, they can be expected to increase the is-HIT amino acid position's resistance to digestion by protease while at the same time, maintaining or improving the requisite biological activity of the protein (these replacing amino acids are the candidate LEADs).**

To select the candidate replacing amino acids for each is-HIT position, PAM250 matrix based analysis was used (FIG7). In one

embodiment, the two highest values in PAM250 matrix, corresponding to the highest occurrence of substitutions between residues ("conservative substitutions" or "accepted point mutations"), were chosen (FIG8). Whenever only a conservative substitution was available for a given high value of the PAM250, the following higher value was selected and the totality of conservative substitutions for this value was considered. The replacement of amino acids that are exposed on the surface by cysteine residues (as shown in FIG8, while replacing Y by H or I) was explicitly avoided, since this change would potentially lead to the formation of intermolecular disulfide bonds.

Thus, based on the nature of the challenging proteases, and on evolutionary considerations as well as protein structural analysis, a strategy was defined for the rational design of human IFN$\alpha$-2b mutants having increased resistance to proteolysis which could produce therapeutic proteins having a longer half-life. By using the algorithm PROTEOL (see, *e.g.*, infobiogen.fr), a list of residues along the IFN$\alpha$-2b sequence was established, which can be recognized as a substrate for different enzymes present in the serum. Because the number of residues in this particular list was high, the 3-dimensional structure of IFN$\alpha$-2b obtained from the NMR structure of IFN$\alpha$-2a (PDB code 1ITF) was used to select only those residues exposed to the solvent. Using this approach, 42 positions were identified, which numbering is that of the mature protein (SEQ ID NO:1): L3, P4, R12, R13, M16, R22, K23, F27, L30, K31, R33, E41, K49, E58, K70, E78, K83, Y89, E96, E107, P109, L110, M111, E113, L117, R120, K121, R125, L128, K131, E132, K133, K134, Y135, P137, M148, R149, E159, L161, R162, K164, and E165. Each of these positions was replaced by amino acid residues, such that they are defined as compatible by the substitution matrix PAM250 while at the same time the replacement amino acids do not generate new sites for proteases.

The list of performed residue substitutions as determined by PAM250 analysis is as follows:

R to H, Q

E to H, Q

K to Q, T

L to V, I

M to I, V

P to A, S

Y to I, H.

**C) Systematically introducing the specific replacing amino acids (candidate LEADs) at every specific is-HIT position to generate a collection containing the corresponding mutant molecules.**

The individual IFN$\alpha$-2b mutants are generated, produced and phenotypically characterized one-by-one, in addressable arrays as set forth in Example 1, such that each mutant molecule contains initially amino acid replacements at only one is-HIT site. LEAD positions were obtained in IFN$\alpha$-2b variants after a screening for protection against proteases, and comparing protease-untreated and protease-treated variant preparations with the corresponding conditions for the wild-type IFN$\alpha$-2b. The percent of residual (anti-viral) activity for the IFN$\alpha$-2b E113H variant after treatment with chymotrypsin, protease mixture, blood lysate or serum was compared to the treated wild-type IFN$\alpha$-2b. Selected IFN$\alpha$-2b LEADs are shown in Table 2.

A top and side view of IFN$\alpha$-2b structure in ribbon representation (obtained from NMR structure of IFN$\alpha$-2b, PDB code 1ITF) depict residues in "space filling" defining (1) the "receptor binding region" as deduced either by "alanine scanning" data and studies by Piehler *et al.*, *J. Biol. Chem.*, 275:40425-40433, 2000, and Roisman *et al.*, *Proc. Natl. Acad. Sci USA*, 98:13231-13236, 2001, and (2) replacing residues (LEADs) for resistance to proteolysis.

**Table 2**

**Selected LEADs of IFN$\alpha$-2b following protease resistance**

| Mutant | SEQ ID No. | Proteolysis protection | IFN antiviral activity |
|---|---|---|---|
| F27V | 83 | Pseudo wt | Pseudo wt |
| R33H | 86 | Pseudo wt | Pseudo wt |
| E41Q | 87 | Increased | Increased |
| E41H | 88 | Pseudo wt | Increased |
| E58Q | 89 | Increased | Pseudo wt |
| E58H | 90 | Increased | Increased |
| E78Q | 92 | Increased | Increased |
| E78H | 93 | Increased | Increased |
| Y89H | 1303 | Pseudo wt | Pseudo wt |
| E107Q | 95 | Increased | Pseudo wt |
| E107H | 96 | Increased | Pseudo wt |
| P109A | 97 | Pseudo wt | Pseudo wt |
| L110V | 98 | Pseudo wt | Pseudo wt |
| M111V | 978 | Pseudo wt | Pseudo wt |
| E113H | 101 | Increased | Pseudo wt |
| L117V | 102 | Increased | Pseudo wt |
| L117I | 103 | Increased | Pseudo wt |
| K121Q | 104 | Increased | Pseudo wt |
| R125H | 106 | Increased | Increased |
| R125Q | 107 | Increased | Increased |
| K133Q | 114 | Increased | Increased |
| E159H | 125 | Increased | Pseudo wt |
| E159Q | 124 | Increased | Pseudo wt |

## EXAMPLE 3

**Stabilization of IFN$\alpha$-2b by Creation of N-Glycosylation Sites**

The creation of N-glycosylation sites on the protein was a second strategy that was used to stabilize IFN$\alpha$-2b Natural human IFN$\alpha$-2b contains a unique *O*-glycosylation site at position 129 (the numbering corresponds to the mature protein; SEQ ID NO:1), however, no *N*-glycosylation sites are found in this sequence. N-glycosylation sites are defined by the N-X-S or N-X-T consensus sequences. Glycosylation has been found to play a role in protein stability. For example, glycosylation

has been found to increase bioavailability via higher metabolic stability and reduced clearance. In order to generate more stable IFN$\alpha$-2b variants, the N-glycosylation consensus sequences indicated above were introduced in the IFN$\alpha$-2b sequence by mutagenesis. Variants of IFN$\alpha$-2b carrying new glycosylation sites were assessed as previously described.

The structure of IFN$\alpha$-2b is characterized by a helix bundle composed of 5 helices (A, B, C, D and E) connected with each other by a series of loops (a large AB loop and three shorter BC, CD, DE loops). The helices are joined together by two disulfide bridges between residues 1/98 and 29/138 of SEQ ID NO:1. The loops are contemplated herein to represent preferential sites for glycosylation given their exposure. Therefore, N-glycosylation sites (N-X-S or N-X-T) were created in each of the loop sequences (Table 3). Selected LEADs and pseudo wild-type IFN$\alpha$-2b mutants after screening for addition of glycosylation sites are shown in Table 4.

**Table 3**
***In silico* HITs for addition of glycosylation sites on IFN$\alpha$-2b**

| Codon No. | SEQ ID No. | N-X-S | SEQ ID No. | N-X-T |
|---|---|---|---|---|
| c2-4 | | D2N/P4S | | D2N/P4T |
| c3-5 | | L3N/Q5S | | L3N/Q5T |
| c4-6 | | P4N/T6S | | P4N/T6T |
| c5-7 | 127 | Q5N/H7S | 128 | Q5N/H7T |
| c6-8 | 129 | T6N/S8S | | T6N/S8T |
| c7-9 | | H7N/L9S | | H7N/L9T |
| c8-10 | 130 | S8N/G10S | 131 | S8N/G10T |
| c9-11 | | L9N/S11S | | L9N/S11T |
| c10-12 | 132 | M21N/R23S | | M21N/R23T |
| c22-24 | | R22N/I24S | | R22N/I24T |
| c23-25 | | R23N/S25S | 133 | R23N/S25T |
| c24-26 | 134 | I24N/L26S | | I24N/L26T |
| c25-27 | 135 | S25N/F27S | 136 | S25N/F27T |
| c26-28 | 137 | L26N/S28S | 138 | L26N/S28T |
| c28-30 | | S28N/L30S | | S28N/L30T |
| c30-32 | 139 | L30N/D32S | | L30N/D32T |

| Codon No. | SEQ ID No. | N-X-S | SEQ ID No. | N-X-T |
|---|---|---|---|---|
| c31-33 | | K31N/R33S | | K31N/R33T |
| c32-34 | | D32N/H34S | | D32N/H34T |
| c33-35 | 140 | R33N/D35S | 141 | R33N/D35T |
| c34-36 | 142 | H34N/F36S | 143 | H34N/F36T |
| c35-37 | 144 | D35N/G37S | | D35N/G37T |
| c36-38 | 145 | F36N/F38S | 146 | F36N/F38T |
| c37-39 | 147 | G37N/P39S | | G37N/P39T |
| c38-40 | 148 | F38N/Q40S | 149 | F38N/Q40T |
| c39-41 | 150 | P39N/E41S | 151 | P39N/E41T |
| c40-42 | 152 | Q40N/E42S | 153 | Q40N/E42T |
| c41-43 | | E41N/F43S | 155 | E41N/F43T |
| c42-44 | | E42N/G44S | | E42N/G44T |
| c43-45 | | F43N/N45S | | F43N/N45T |
| c44-46 | 156 | G44N/Q46S | 157 | G44N/Q46T |
| c45-47 | 158 | N45N/F47S | 159 | N45N/F47T |
| c46-48 | 160 | Q46N/Q48S | 161 | Q46N/Q48T |
| c47-49 | 162 | F47N/K49S | 163 | F47N/K49T |
| c48-50 | | Q48N/A50S | | Q48N/A50T |
| c49-51 | 164 | K49N/E51S | | K49N/E51T |
| c50-52 | | A50N/T52S | | A50N/T52T |
| c68-70 | | S68N/K70S | | S68N/K70T |
| c70-72 | | K70N/S72S | | K70N/S72T |
| c75-77 | 165 | A75N/D77S | | A75N/D77T |
| c77-79 | | D77N/T79S | | D77N/T79T |
| C100-102 | 166 | I100N/G102S | 167 | I100N/G102T |
| C101-103 | | Q101N/V103S | | Q101N/V103T |
| C102-104 | | G102N/G104S | | G102N/G104T |
| C103-105 | 168 | V103N/V105S | 169 | V103N/V105T |
| C104-106 | | G104N/T106S | 170 | G104N/T106T |
| C105-107 | 171 | V105N/E107S | | V105N/E107T |
| C10--108 | 172 | T106N/T108S | 173 | T106N/T108T |
| C107-109 | 174 | E107N/P109S | 175 | E107N/P109T |
| C108-110 | | T108N/I110S | | T108N/I110T |
| C134-136 | | K134N/S136S | 176 | K134N/S136T |
| C154-156 | | S154N/N156S | | S154N/N156T |
| C155-157 | | T155N/L157S | | T155N/L157T |

| Codon No. | SEQ ID No. | N-X-S | SEQ ID No. | N-X-T |
|---|---|---|---|---|
| C156-158 | | N156N/Q158S | | N156N/Q158T |
| C157-159 | 177 | L157N/E159S | 178 | L157N/E159T |
| C158-160 | | Q158N/S160S | 179 | Q158N/S160T |
| C159-161 | 180 | E159N/L161S | 181 | E159N/L161T |
| C160-162 | | S160N/R162S | | S160N/R162T |
| C161-163 | | L161N/S163S | | L161N/S163T |
| C162-164 | | R162N/K164S | | R162N/K164T |
| C163-165 | | S163N/E165S | | S163N/E165T |

**Table 4**

**Selected LEADs and pseudo wild-type IFN$\alpha$-2b mutants after screening for addition of glycosylation sites**

| Mutant | SEQ ID No. | Proteolysis protection | IFN antiviral activity |
|---|---|---|---|
| Q5N/H7S | 127 | Increased | Pseudo wt |
| Q5N/H7T | 128 | ND* | ND |
| P39N/E41S | 150 | Increased | Pseudo wt |
| P39N/E41T | 151 | Increased | Pseudo wt |
| Q40N/E42S | 152 | Increased | Pseudo wt |
| Q40N/E42T | 153 | Increased | Pseudo wt |
| E41N/F43S | 154 | Increased | Pseudo wt |
| E41N/F43T | 155 | Increased | Pseudo wt |
| F43N/N45S | | Increased | Pseudo wt |
| F43N/N45T | | ND | ND |
| G44N/Q46S | 156 | ND | ND |
| G44N/Q46T | 157 | Increased | Pseudo wt |
| N45N/F47S | 158 | Increased | Pseudo wt |
| N45N/F47T | 159 | Increased | Pseudo wt |
| Q46N/Q48S | 160 | Increased | Pseudo wt |
| Q46N/Q48T | 161 | ND | ND |
| F47N/K49S | 162 | Increased | Pseudo wt |
| F47N/K49T | 163 | Increased | Pseudo wt |
| I100N/G102S | 166 | Pseudo wt | Increased |
| I100N/G102T | 167 | Pseudo wt | Increased |
| V105N/E107S | 171 | Pseudo wt | Increased |
| V105N/E107T | | Pseudo wt | Increased |
| T106N/T108S | 172 | Pseudo wt | Increased |
| T106N/T108T | 173 | Pseudo wt | Increased |
| E107N/P109S | 174 | Pseudo wt | Increased |
| E107N/P109T | 175 | Pseudo wt | Increased |
| L157N/E159S | 177 | Pseudo wt | Increased |
| L157N/E159T | 178 | Pseudo wt | Increased |
| E159N/L161S | 180 | Pseudo wt | Increased |
| E159N/L161T | 181 | Pseudo wt | Increased |

*ND, not determined

## Example 4

## Redesign of Interferon α-2b Proteins

The use of the protein redesign approach provided herein permits the generation of proteins such that they maintain requisite levels and types of biological activity compared to the native protein while their underlying amino acid sequences have been significantly changed by amino acid replacement. To first identify those amino acid positions on the IFNα-2b protein that are involved or not involved IFNα-2b protein activity, such as binding activity of IFNα-2b to its receptor, an Ala-scan was performed on the IFNα-2b sequence. For this purpose, each amino acid in the IFNα-2b protein sequence was individually changed into Alanine. Any other amino acid, particularly another amino acid that has a neutral effect on structure, such as Gly or Ser, also can be used. Each resulting mutant IFNα-2b protein was then expressed and the antiviral activity of the individual mutants was assayed. The particular amino acid positions that are sensitive to replacement by Ala, referred to herein as HITs would in principle not be suitable targets for amino acid replacement to increase protein stability, because of their involvement in the activity of the molecule. For the Ala-scanning, the biological activity measured for the IFNα-2b molecules was: *i*) their capacity to inhibit virus replication when added to permissive cells previously infected with the appropriate virus and, *ii*) their capacity to stimulate cell proliferation when added to the appropriate cells. The relative activity of each individual mutant compared to the native protein was assayed. HITS are those mutants that produce a decrease in the activity of the protein (e.g., in this example, all the mutants with activities below about 30% of the native activity).

In addition, to identify the HIT positions, the Alanine-scan was used to identify the amino acid residues on IFNα-2b that when replaced with alanine lead to a 'pseudo-wild type' activity, i.e., those that can be replaced by alanine without leading to a decrease in biological activity.

A collection of mutant molecules was generated and phenotypically characterized such that IFN$\alpha$-2b proteins with amino acid sequences different from the native ones but that still elicit the same level and type of activity as the native protein were selected. HITs and pseudo wild-type amino acid positions are shown in Table 5.

**Table 5**

**HITs and pseudo wild-type positions to IFN$\alpha$-2b redesign**

| Mutants | SEQ ID No. | HITs (viral activity) | Pseudo wt (viral activity) |
|---|---|---|---|
| D2A | 2 | Decreased | |
| P4A | 3 | | Pseudo wt |
| Q5A | 4 | | Pseudo wt |
| T6A | 5 | | Pseudo wt |
| H7A | 6 | Decreased | |
| S8A | 7 | Decreased | |
| L9A | 8 | | Pseudo wt |
| G10A | 9 | | Pseudo wt |
| S11A | 10 | Decreased | |
| R12A | 11 | Decreased | |
| R13A | 12 | Decreased | |
| T14A | 13 | Decreased | |
| L15A | 14 | Decreased | |
| M16A | 15 | Decreased | |
| L17A | 16 | | Pseudo wt |
| Q20A | 17 | | Pseudo wt |
| R23A | 18 | Decreased | |
| I24A | 19 | | Pseudo wt |
| S25A | 20 | | Pseudo wt |
| L26A | 21 | Decreased | |
| S28A | 22 | Decreased | |
| C29A | 23 | Decreased | |
| L30A | 24 | Decreased | |
| K31A | 25 | Decreased | |
| D32A | 26 | Decreased | |
| R33A | 27 | Decreased | |
| D35A | 28 | | Pseudo wt |
| G37A | 29 | | Pseudo wt |
| G39A | 30 | | Pseudo wt |

| Mutants | SEQ ID No. | HITs (viral activity) | Pseudo wt (viral activity) |
|---|---|---|---|
| E41A | 31 | | Pseudo wt |
| E42 | 32 | | Pseudo wt |
| F43A | 33 | Decreased | |
| N45A | 34 | Decreased | |
| F47A | 35 | Decreased | |
| E51A | 36 | | Pseudo wt |
| T52A | 37 | | Pseudo wt |
| I53A | 38 | Decreased | |
| P54A | 39 | | Pseudo wt |
| V55A | 40 | | Pseudo wt |
| L56A | 41 | | Pseudo wt |
| H57A | 42 | | Pseudo wt |
| E58A | 43 | | Pseudo wt |
| M59A | 44 | Decreased | |
| I60A | 45 | | Pseudo wt |
| I63A | 46 | | Pseudo wt |
| F64A | 47 | | Pseudo wt |
| N65A | 48 | | Pseudo wt |
| L66A | 49 | Decreased | |
| F67A | 50 | Decreased | |
| T69A | 51 | Decreased | |
| K70A | 52 | Decreased | |
| D71A | 53 | Decreased | |
| S72A | 54 | Decreased | |
| W76A | 55 | | Pseudo wt |
| D77A | 56 | | Pseudo wt |
| E78A | 57 | | Pseudo wt |
| L81A | 58 | | Pseudo wt |
| D82A | 59 | Decreased | |
| K83A | 60 | Decreased | |
| F84A | 61 | Decreased | |
| Y85A | 62 | | Pseudo wt |
| Y89A | 63 | | Pseudo wt |
| Q90A | 64 | | Pseudo wt |
| Q91 | 65 | Decreased | |
| N93A | 66 | Decreased | |

| Mutants | SEQ ID No. | HITs (viral activity) | Pseudo wt (viral activity) |
|---|---|---|---|
| D94A | 67 | Decreased | |
| C98A | 68 | Decreased | |
| V99A | 69 | Decreased | |
| Q101A | 207 | Decreased | |
| G104A | 70 | | Pseudo wt |
| L110A | 71 | | Pseudo wt |
| S115A | 72 | | Pseudo wt |
| Y122A | 73 | Decreased | |
| W140A | 74 | Decreased | |
| E146A | 75 | | Pseudo wt |

## EXAMPLE 5

**Super LEADS of Interferon *α*-2b Protein by Additive Directional Mutagenesis**

The use of an additive directional mutagenesis approach provided a method for the assembly of multiple mutations previously present on the individual LEAD molecules in a single mutant protein thereby generating super-LEAD mutant proteins. In this method, a collection of nucleic acid molecules encoding a library of new mutant molecules is generated, tested and phenotypically characterized one-by-one in addressable arrays. Super-LEAD mutant molecules are such that each molecule contains a variable number and type of LEAD mutations

Using the LEADs obtained in Example 2, six series of mutant molecules were generated with more than one mutation per molecule as shown in Table 6. Some SuperLEAD mutant molecules were phenotypically characterized and the results are shown in Table 7. As shown in the table not all SuperLEADS have improved activity compared with the original Leads; some showed decreased activity of some type.

**Table 6**

Schema of LEADs position for SuperLEADS generation

Series 1

m1 = E41H

m1 + m2 = E41H + Y89H

Series 2

m1 = E58Q

m1 + m2 = E58Q + F27V

Series 3

m1 = R125H

m1 + m2 = R125H + M111V

Series 4

m1 = E159H

m1 + m2 = E159H + Y89H

Series 5

m1 = K121Q

m1 + m2 = K121Q + P109A

m1 + m2 + m3 = K121Q + P109A + K133Q

Series 6

m1 = E78H

m1 + m2 = E78H + R33H

m1 + m2 + m3 = E78H + R33H + E58H

m1 + m2 + m3 + m4 = E78H + R33H + E58H + L110V

**Table 7**

**SuperLEADs of IFN$\alpha$-2b multiple mutants**

| Mutant | SEQ ID No. | Proteolysis protection | IFN antiviral activity |
|---|---|---|---|
| E41H | 88 | Pseudo wt | Increased |
| Y89H | 1303 | Pseudo wt | Pseudo wt |
| E41H/Y89H/ N45D** | 979 | Increased | Increased |
| E58Q | 89 | Increased | Pseudo wt |
| F27V | 83 | Pseudo wt | Pseudo wt |
| E58Q/F27V | 981 | Increased | Pseudo wt |
| R125H | 106 | Increased | Increased |
| M111V | 978 | Pseudo wt | Pseudo wt |
| R125H/M111V | 986 | Increased | Increased |
| E159H | 125 | | |
| Y89H | 1303 | | |
| E159H/Y89H | 987 | | |

| Mutant | SEQ ID No. | Proteolysis prot ction | IFN antiviral activity |
|---|---|---|---|
| K121Q | 104 | Increased | Pseudo wt |
| P109A | 97 | Pseudo wt | Pseudo wt |
| K133Q | 114 | Increased | Increased |
| K121Q/P109A | 983 | Increased | Pseudo wt |
| K121Q/P109A / K133Q / G102R | 984 | Increased | Increased |
| E78H | 93 | Increased | Increased |
| R33H | 86 | Pseudo wt | Pseudo wt |
| E58H | 89 | Increased | Increased |
| L110V | 98 | Pseudo wt | Pseudo wt |
| E78H/R33H/ E58H/L110V | 982 | Decreased | Decreased |

Four mutants with mutations to additional those selected by the rational mutagenesis were generated in the *E. coli* MutS strain and were detected by sequencing. The mutants were the following: E41Q/ D94G SEQ.ID No. 199; L117V/ A139G SEQ.ID No. 204; E41H/ Y89H/ N45D SEQ.ID No. 198; and K121Q/ P109A/ K133Q/ G102R SEQ.ID No. 204.

## EXAMPLE 6

**Cloning of IFN β in pNAUT, a mammalian cell expression plasmid**

The cDNA encoding IFN β (see, SEQ ID No. 196) was cloned into a mammalian expression vector, prior to the generation of the selected mutations (see, Figures 6(O)-6(S) and 8(A). A collected of predesigned, targeted mutants was then generated such that each individual mutant was created and processed individually, physically separated form each other and in addressable arrays. The mammalian expression vector pSSV9 CMV 0.3 pA (see, Example 1) was engineered as follows:

The pSSV9 CMV 0.3 pA was cut by *Pvu*II and religated (this step gets rid of the ITR functions), prior to the introduction of a new *Eco*RI restriction site by Quickchange mutagenesis (Stratagene). The oligonucleotides sequences used, follow:

EcoRI forward primer: 5'-GCCTGTATGATTTATTGGATGTTGGAATTCCCTGAT-GCGGTATTTTCTCCTTACG-3' (SEQ ID NO: 218)

EcoRI reverse prime: 5'-CGTAAGGAGAAAATACCGCATCAGG-GAATTCCAACATCCAATAAATCATACAGGC-3' (SEQ ID NO: 219)

The construct sequence was confirmed by using the following oligonucleotides:

Seq ClaI forward primer: 5'-CTGATTATCAACCGGGGTACATATGATTGACATGC-3' (SEQ ID NO: 220)

Seq XmnI reverse primer 5'-TACGGGATAATACCGCGCCACATAGCAGAAC-3' (SEQ ID NO: 221).

Then, the *Xmn*I-*Cla*I fragment containing the newly introduced *Eco*RI site was cloned into pSSV9 CMV 0.3 pA to replace the corresponding wild-type fragment and produce construct pSSV9-2EcoRI.

The IFN $\beta$-cDNA was obtained from the pIFN$\beta$1 (ATCC) construct. The sequence of the IFN $\beta$-cDNA was confirmed by sequencing using the primers below:

Seq forward primer: 5'-CCTGATGAAGGAGGACTC-3' (SEQ ID NO:222)

Seq reverse primer: 5'-CCAAGCAGCAGATGAGTC-3' (SEQ ID NO:223).

The verified IFN $\beta$-encoding cDNA first was cloned into the pTOPO-TA vector (Invitrogen). After checking of the cDNA sequence by automatic DNA sequencing, the *Hin*dIII-*Xba*I fragment containing the IFN cDNA was subcloned into the corresponding sites of pSSV9-2EcoRI, leading to the construct pAAV-EcoRI-IFNbeta (pNB-AAV-IFN beta) Finally the fragment Pvu II of plasmid pNB-AAV-IFN beta was subcloned in PvuII site of pUC 18 leading the final construct pUC-CMVIFNbetapA called pNAUT-IFNbeta.

**Production and normalization of IFN$\beta$ in mammalian cells**

IFN $\beta$ was produced in CHO Chinese Hamster Ovarian cells (obtained from ATCC), using Dubelcco's modified Eagle's medium supplemented with glucose (4.5 g/L; Gibco-BRL) and fetal bovine serum (5 %, Hyclone). Cells were transiently transfected as follows: 0.6 x $10^5$ cells were seeded into 6 well plates and grown for 24 h before transfection. Confluent cells at about 70%, were supplemented with 1.0 $\mu$g of plasmid (from the library of IFN $\beta$ mutants) by lipofectamine plus

reagent (Invitrogen). After gently shaking, cells were incubated for 24 h with 1 ml of culture medium supplemented with 1 % of serum. IFN $\beta$ was obtained from culture supernatants 24 h after transfection and stored in aliquots at −80 °C until use.

Preparations of IFN $\beta$ produced from transfected cells were screened following sequential biological assays as follows. Normalization of IFN $\beta$ concentration from culture supernatants was performed by ELISA. IFN $\beta$ concentrations from wild type, and mutants samples were estimated by using an international reference standard provided by the NIBSC, UK.

**Screening and in vitro charaterization of IFN $\beta$ mutants**

Two activities were measured directly on IFN samples: antiviral and antiproliferation activities. Dose (concentration) - response (activity) experiments for antiviral or antiproliferation activity allowed for the calculation of the 'potency' for antiviral and antiproliferation activities, respectively. Antiviral and antiproliferation activities also were measured after incubation with proteolytic samples such as specific proteases, mixtures of selected proteases, human serum or human blood. Assessment of activity following incubation with proteolytic samples allowed to determine the residual (antiviral or antiproliferation) activity an.d the respective kinetics of half-life upon exposure to proteases

**Antiviral activity - measured by Cytopathic Effects (CPE)**

Antiviral activity of IFN $\beta$ was determined by the capacity of the cytokine to protect Hela cells against EMC (mouse encephalomyocarditis) virus-induced cytopathic effects. The day before, Hela cells ($2x10^5$ cells/ml) were seeded in flat-bottomed 96-well plates containing 100 $\mu$l/well of Dulbecco's MEM-Glutamaxl-sodium pyruvate medium supplemented with 5% SVF and 0.2% of gentamicin. Cells were growth at 37°C in an atmosphere of 5% $CO_2$ for 24 hours

Two-fold serial dilutions of interferon samples were made with MEM complete media into 96-Deep-Well plates with final concentration ranging

from 1600 to 0.6 pg/ml. The medium was aspirated from each well and 100 µl of interferon dilutions were added to Hela cells. Each interferon sample dilution was assessed in triplicate. The two last rows of the plates were filled with 100 µl of medium without interferon dilution samples in order to serve as controls for cells with and without virus.

After 24 hours of growth, a 1/1000 EMC virus dilution solution was placed in each well, except for the cell control row. Plates were returned to the $CO_2$ incubator for 48 hours. Then, the medium was aspirated and the cells were stained for 1 hour with 100 µl of Blue staining solutio to determine the proportion of intact cells. Plates were washed in a distilled water bath. The cell bound dye was extracted using 100 µl of ethylene-glycol mono-ethyl-ether (Sigma). The absorbance of the dye was measured using an Elisa plate reader (Spectramax). The antiviral activity of IFN β samples (expressed as number of IU/mg of proteins) was determined as the concentration needed for 50% protection of the cells against EMC virus-induced cytopathic effects. For proteolysis experiments, each point of the kinetic was assessed at 800 and 400 pg/ml in triplicate.

**Anti-proliferative activity**

Anti-proliferative activity of IFN β was determined by assessing the capacity of the cytokine to inhibit proliferation of Daudi cells. Daudi cells ($1x10^4$ cells) were seeded in flat-bottomed 96-well plates containing 50µl/well of RPMI 1640 medium supplemented with 10% SVF, 1X glutamine and 1ml of gentamicin. No cell was added to the last row ("H" row) of the flat-bottomed 96-well plates in order to evaluate background absorbance of culture medium.

At the same time, two-fold serial dilutions of interferon samples were made with RPMI 1640 complete media into 96-Deep-Well plates with final concentration ranging from 6000 to 2.9 pg/ml. Interferon dilutions (50µl) were added to each well containing 50µl of Daudi cells. The total volume in each well should now be 100µl. Each interferon

sample dilution was assessed in triplicate. Each well of the "G" row of the plates was filled with 50$\mu$l of RPMI 1640 complete media in order to be used as positive control. The plates were incubated for 72 hours at 37°C in a humidified, 5% CO2 atmosphere.

After 72 hours of growth, 20 $\mu$l of Cell titer 96 Aqueous one solution reagent (Promega) was added to each well and incubated 1H30 at 37°C in an atmosphere of 5% $CO_2$. To measure the amount of colored soluble formazan produced by cellular reduction of the MTS, the absorbance of the dye was measured using an Elisa plate reader (spectramax) at 490nm.

The corrected absorbances ("H" row background value subtracted) obtained at 490nm were plotted versus concentration of cytokine. The ED50 value was calculated by determining the X-axis value corresponding to one-half the difference between the maximum and minimum absorbance values. (ED50 = the concentration of cytokine necessary to give one-half the maximum response).

**Treatment of IFN $\beta$ with proteolytic preparations**

Mutants were treated with proteases in order to identify resistant molecules. The resistance of the mutant IFN $\beta$molecules compared to wild-type IFN $\beta$ against enzymatic cleavage (120 min, 25 °C) by a mixture of proteases (containing 1.5 pg of each of the following proteases (1% wt/wt, Sigma): $\alpha$-chymotrypsin, carboxypeptidase, endoproteinase Arg-C, endoproteinase Asp-N, endoproteinase Glu-C, endoproteinase Lys-C, and trypsin) was determined. At the end of the incubation time, 10 $\mu$l of anti-proteases complete, mini EDTA free, Roche (one tablet was dissolved in 10 ml of DMEM and then diluted to 1/1000) was added to each reaction in order to inhibit protease activity. Treated samples were then used to determine residual antiviral or antiproliferation activities.

**Protease resistance - Kinetic analysis**

The percent of residual IFN $\beta$ activity over time of exposure to proteases was evaluated by a kinetic study using 1.5 pg of protease mixture. Incubation times were: 0 h, 0.5 h, 2 h, 4 h, 8 h, 12 h, 24 h and 48 h. Briefly, 20 $\mu$l of each proteolytic sample (proteases, serum, bnlood) was added to 100 $\mu$l of IFN $\beta$ at 400 and 800 pg/ml and incubated for variable times, as indicated. At the appropriate time points, 10 $\mu$l of anti-proteases mixture, mini EDTA free, Roche (one tablet was dissolved in 10 ml of DMEM and then diluted to 1/500) was added to each well in order to stop proteolysis reactions. Biological activity assays were then performed as described for each sample in order to determine the residual activity at each time point.

**Performance**

The various biological activities, protease resistance and potency of each individual mutant were analyzed using a mathematical model and algorithm (NautScan™; Fr. Patent No. 9915884; see, also published International PCT application No. WO 01/44809 based on PCT n° PCT/FR00/03503). Data was processed using a Hill equation-based model that uses key feature indicators of the performance of each individual mutant. Mutants were ranked based on the values of their individual performance and those on the top of the ranking list were selected as leads.

Using the 2D-scanning and 3D-scanning methods described above in addition to the 3-dimensional structure of IFN$\beta$, the following amino acid target positions were identified as is-HITs on IFN$\beta$, which numbering is that of the mature protein (SEQ ID NO:196):

**By 3D-scanning** (see, SEQ ID Nos. 234-289, 989-1015): D by Q at position 39, D by H at position 39, D by G at position 39, E by Q at position 42, E by H at position 42, K by Q at position 45, K by T at position 45, K by S at position 45, K by H at position 45, L by V at position 47, L by I at position 47, L by T at position 47, L by Q at position 47, L by H at position 47, L by A at position 47, K by Q at

position 52, K by T at position 52, K by S at position 52, K by H at position 52, F by I at position 67, F by V at position 67, R by H at position 71, R by Q at position 71, D by H at position 73, D by G at position 73, D by Q at position 73, E by Q at position 81, E by H at position 81, E by Q at position 107, E by H at position 107, K by Q at position 108, K by T at position 108, K by S at position 108, K by H at position 108, E by Q at position 109, E by H at position 109, D by Q at position 110, D by H at position 110, D by G at position 110, F by I at position 111, F by V at position 111, R by H at position 113, R by Q at position 113, L by V at position 116, L by I at position 116, L by T at position 116, L by Q at position 116, L by H at position 116, L by A at position 116, L by V at position 120, L by I at position 120, L by T at position 120, L by Q at position 120, L by H at position 120, L by A at position 120, K by Q at position 123, K by T at position 123, K by S at position 123, K by H at position 123, R by H at position 124,, R by Q at position 124, R by H at position 128, R by Q at position 128, L by V at position 130, L by I at position 130, L by T at position 130, L by Q at position 130, L by H at position 130, L by A at position 130, K by Q at position 134, K by T at position 134, K by S at position 134, K by H at position 134, K by Q at position 136, K by T at position 136, K by S at position 136,, K by H at position 136, E by Q at position 137, E by H at position 137, Y by H at position 163, Y by I at position 163I, R by H at position 165, R by Q at position 165.

**By 2D-scanning** (see, SEQ ID Nos.1016-1302, and table above): M by V at position 1, M by I at position 1, M by T at position 1, M by Q at position 1, M by A at position 1, L by V at position 5, L by I at position 5, L by T at position 5, L by Q at position 5, L by H at position 5, L by A at position 5, F by I at position 8, F by V at position 8, L by V at position 9, L by I at position 9, L by T at position 9, L by Q at position 9, L by H at position 9, L by A at position 9, R by H at position 11, R by Q at position 11, F by I at position 15, F by V at position 15, K by Q at position 19, K

by T at position 19, K by S at position 19, K by H at position 19, W by S at position 22, W by H at position 22, N by H at position 25, N by S at position 25, N by Q at position 25, R by H position 27, R by Q position 27, L by V at position 28, L by I at position 28, L by T at position 28, L by Q at position 28, L by H at position 28, L by A at position 28, E by Q at position 29, E by H at position 29, Y by H at position 30, Y by I at position 30, L by V at position 32, L by I at position 32, L by T at position 32, L by Q at position 32, L by H at position 32, L by A at position 32, K by Q at position 33, K by T at position 33, K by S at position 33, K by H at position 33, R by H at position 35, R by Q at position 35, M by V at position 36, M by I at position 36, M by T at position 36, M by Q at position 36, M by A at position 36, D by Q at position 39, D by H at position 39, D by G at position 39, E by Q at position 42, E by H at position 42, K by Q at position 45, K by T at position 45, K by S at position 45, K by H at position 45, L by V at position 47, L by I at position 47, L by T at position 47, L by, Q at position 47, L by H at position 47, L by A at position 47, K by Q at position 52, K by T at position 52, K by S at position 52, K by H at position 52, F by I at position 67, F by V at position 67, R by H at position 71, R by Q at position 71, D by Q at position 73, D by H at position 73, D by G at position 73, E by Q at position 81, E by H at position 81, E by Q at position 85, E by H at position 85, Y by H at position 92, Y by I at position 92, K by Q at position 99, K by T at position 99, K by S at position 99, K by H at position 99, E by Q at position 103, E by H at position 103, E by Q at position 104, E by H at position 104, K by Q at position 105, K by T at position 105, K by S at position 105, K by H at position 105, E by Q at position 107, E by H at position 107, K by Q at position 108, K by T at position 108, K by S at position 108, K by H at position 108, E by Q at position 109, E by H at position 109, D by Q at position 110, D by H at position 110, D by G at position 110, F by I at position 111, F by V at position 111, R by H at

position 113, R by Q at position 113, L by V at position 116, L by I at position 116, L by T at position 116, L by Q at position 116, L by H at position 116, L by A at position 116, L by V at position 120, L by I at position 120, L by T at position 120, L by Q at position 120, L by H at position 120, L by A at position 120, K by Q at position 123, K by T at position 123, K by S at position 123, K by H at position 123, R by H at position 124, R by Q at position 124, R by H at position 128, R by Q at position 128, L by V at position 130, L by I at position 130, L by T at position 130, L by Q at position 130, L by H at position 130, L by A at position 130, K by Q at position 134, K by T at position 134, K by S at position 134, K by H at position 134, K by Q at position 136, K by T at position 136, K by S at position 136, K by H at position 136, E by Q at position 137, E by H at position 137, Y by H at position 138, Y by I at position 138, R by H at position 152, R by Q at position 152, Y by H at position 155, Y by I at position 155, R by H at position 159, R by Q at position 159, Y by H at position 163, Y by I at position 163, R by H at position 165, R by Q at position 165, M by D at position 1, M by E at position 1, M by K at position 1, M by N at position 1, M by R at position 1, M by S at position 1, L by D at position 5, L by E at position 5, L by K at position 5, L by N at position 5, L by R at position 5, L by S at position 5, L by D at position 6, L by E at position 6, L by K at position 6, L by N at position 6, L by R at position 6, L by S at position 6, L by Q at position 6, L by T at position 6, F by E at position 8, F by K at position 8, F by R at position 8, F by D at position 8, L by D at position 9, L by E at position 9, L by K at position 9, L by N at position 9, L by R at position 9, L by S at position 9, Q by D at position 10, Q by E at position 10, Q by K at position 10, Q by N at position 10, Q by R at position 10, Q by S at position 10, Q by T at position 10, S by D at position 12, S by E at position 12, S by K at position 12, S by R at position 12, S by D at position 13, S by E at position 13, S by K at position 13, S by R at position 13, S by N at position 13, S by Q at position 13, S by T at

position 13, N by D at position 14, N by E at position 14, N by K at position 14, N by Q at position 14, N by R at position 14, N by S at position 14, N by T at position 14, F by D at position 15, F by E at position 15, F by K at position 15, F by R at position 15, Q by D at position 16, Q by E at position 16, Q by K at position 16, Q by N at position 16, Q by R at position 16, Q by S at position 16, Q by T at position 16, C by D at position 17, C by E at position 17, C by K at position 17, C by N at position 17, C by Q at position 17, C by R at position 17, C by S at position 17, C by T at position 17, L by N at position 20, L by Q at position 20, L by R at position 20, L by S at position 20, L by T at position 20, L by D at position 20, L by E at position 20, L by K at position 20, W by D at position 22, W by E at position 22, W by K at position 22, W by R at position 22, Q by D at position 23, Q by E at position 23, Q by K at position 23, Q by R at position 23, L by D at position 24, L by E at position 24, L by K at position 24, L by R at position 24, W by D at position 79, W by E at position 79, W by K at position 79, W by R at position 79, N by D at position 80, N by E at position 80, N by K at position 80, N by R at position 80, T by D at position 82, T by E at position 82, T by K at position 82, T by R at position 82, I by D at position 83, I by E at position 83, I by K at position 83, I by R at position 83, I by N at position 83, I by Q at position 83, I by S at position 83, I by T at position 83, N by D at position 86, N by E at position 86, N by K at position 86, N by R at position 86, N by Q at position 86, N by S at position 86, N by T at position 86, L by D at position 87, L by E at position 87, L by K at position 87, L by R at position 87, L by N at position 87, L by Q at position 87, L by S at position 87, L by T at position 87, A by D at position 89, A by E at position 89, A by K at position 89, A by R at position 89, N by D at position 90, N by E at position 90, N by K at position 90, N by Q at position 90, N by R at position 90, N by S at position 90, N by T at position 90, V by D at position 91, V by E at

position 91, V by K at position 91, V by N at position 91, V by Q at position 91, V by R at position 91, V by S at position 91, V by T at position 91, Q by D at position 94, Q by E at position 94, Q by Q at position 94, Q by N at position 94, Q by R at position 94, Q by S at position 94, Q by T at position 94, I by D at position 95, I by E at position 95, I by K at position 95, I by N at position 95, I by Q at position 95, I by R at position 95, I by S at position 95, I by T at position 95, H by D at position 97, H by E at position 97, H by K at position 97, H by N at position 97, H by Q at position 97, H by R at position 97, H by S at position 97, H by T at position 97, L by D at position 98, L by E at position 98, L by K at position 98, L by N at position 98, L by Q at position 98, L by R at position 98, L by S at position 98, L by T at position 98, V by D at position 101, V by E at position 101, V by K at position 101, V by N at position 101, V by Q at position 101, V by R at position 101, V by S at position 101, V by T at position 101, M by C at position 1, L by C at position 6, Q by C at position 10, S by C at position 13, Q by C at position 16, L by C at position 17, V by C at position 101, L by C at position 98, H by C at position 97, Q by C at position 94, V by C at position 91, N by C at position 90.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.